

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 15:49:28 ; Search time 40 Seconds

(without alignments)  
738.464 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALGQVRNTELRKEKSR.....TESSLPSWVLMALNRKNCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	802	49.0	667	2 JC7771	hypoxia inducible
2	503	30.7	811	2 JC7619	hypoxia-inducible
3	495	30.3	826	2 I18972	hypoxia-inducible
4	492	30.1	810	2 JC4837	hypoxia-inducible
5	485	29.6	813	2 JC5809	hypoxia-inducible
6	281	17.2	1505	2 JC4851	hypoxia-inducible
7	256	15.6	248	2 A58520	single-minded gene
8	202	12.3	655	2 A29945	neurogenesis regul
9	169.5	10.4	823	2 T11943	hypothetical prote
10	169.5	10.4	825	2 T21944	hypothetical prote
11	142	8.7	805	2 JC7635	aryl hydrocarbon r
12	138.5	8.5	392	2 JC7633	aryl hydrocarbon r
13	138.5	8.5	791	2 A56241	aryl hydrocarbon r
14	136.5	8.3	776	2 A55448	Ah receptor nuclea
15	136.5	8.3	789	2 I19550	aryl hydrocarbon r
16	129.5	7.9	358	2 S58376	aryl hydrocarbon r
17	129.5	7.9	451	2 T42397	aryl hydrocarbon r
18	129	7.9	453	2 T19440	hypothetical prote
19	128	7.8	56	2 A11647	neurogenesis regul
20	125.5	7.7	626	2 JC5405	brain and muscle A
21	123.5	7.5	626	2 JE0270	brain and muscle A
22	118	7.2	701	2 JC8065	Arnt-like PAS prot
23	117.5	7.2	834	2 JC7993	aryl hydrocarbon r
24	106	6.5	848	2 S59514	aryl hydrocarbon r
25	104.5	6.4	1059	2 T30557	aryl hydrocarbon r
26	103.5	6.3	258	2 JC5407	aryl hydrocarbon r
27	103	6.3	846	2 JC7721	brain and muscle A
28	103	6.3	1079	2 A10972	probable DNA polym
29	102.5	6.3	271	2 PC4288	brain and muscle A

30	102	6.2	716	2 T09462	juvenile hormone r
31	98	6.0	1058	2 T30556	aryl hydrocarbon r
32	97.5	6.0	326	2 S71755	transcription fact
33	97	5.9	527	2 A75339	hypothetical prote
34	95.5	5.8	920	2 JC7313	aryl hydrocarbon r
35	94.5	5.8	300	2 T24292	hypothetical prote
36	94.5	5.8	1151	2 T24541	hypothetical prote
37	93.5	5.7	634	2 T00359	hypothetical prote
38	92.5	5.7	1751	2 T09394	gag-pro-pol polypr
39	90	5.5	743	2 G83726	asimilatory nitra
40	90	5.5	1920	2 T13893	gene hindsight pro
41	89.5	5.5	1891	2 T13594	hypothetical prote
42	89	5.4	853	2 S58375	aryl hydrocarbon r
43	89	5.4	3133	2 S52093	hemocytin - silkw
44	87.5	5.3	647	2 S06450	steroid hormone re
45	87.5	5.3	805	2 A46266	aryl hydrocarbon r

ALIGNMENTS

RESULT 1

JC7771  
hypoxia inducible factor-3 alpha - human

C/Species: Homo sapiens (man)

C/Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004

C/Accession: JC7771

R/Hara, S.; Hamada, J.; Kobayashi, C.; Kondo, Y.; Imura, N.

Biochem. Biophys. Res. Commun. 287, 808-813, 2001

A/Title: Expression and characterization of hypoxia-inducible factor (HIF)-3alpha in huma

A/Reference number: JC7771; PMID:11573933

A/Contents: Kidney

A/Accession: JC7771

A/Molecule type: mRNA

A/Residues: 1-667 <HAR>

A/Cross-references: UNIPROT:Q9Y2N7; DBJ:AB054067

A/Comment: This protein is a heterodimeric transcription factor that belongs to the basic

lived in the regulation of hypoxia-inducible gene expression in human kidney.

C/Genetics:

A/Gene: hif-3alpha

A/Map position: 19

C/Keywords: Kidney

Query Match 49.0%; Score 802; DB 2; Length 667;  
Best Local Similarity 68.0%; Pred. No. 4.9e-62;  
Matches 172; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

QY	9	RSNTELRKEKSRDAARSRSSQETEVLYQLAHTLPFARGVSAHLDKASIMELTTSYLMHR	68
DB	7	RSTTELRKEKSRDAARSRSSQETEVLYQLAHTLPFARGVSAHLDKASIMELTTSYLMHR	66
QY	69	LCAAGGKRGATGRLPBGSGFRHGRGRGLPVGKQ-----QARGQSV	117
DB	67	LCAAG-----EWNQVAGGEPDADACYLALRGFVWLTAEQGMAY	106
QY	118	DLSSSLIHNPTPGTWS-DELIGHSIFDTHPCDDELDALTPRNLSSKKLEAPTR	176
DB	107	-----LSENVSKHLGSLQELIGHSTFDTHPCDDELDALTPQTLSSRKVEAPTR	160
QY	177	HPSLRMKSTLTSRGRTNLNKAATWKLHCSGHRAYKPPAQTSPAGSPSEPLQCLVLI	236
DB	161	CFSLRMKSTLTSRGRTNLNKAATWKLHCSGHRAYKPPAQTSPAGSPSEPLQCLVLI	220
QY	237	CEAIPQLPFRHGA 249	
DB	221	CEAIP---HPGS 229	

RESULT 2

JC7619  
hypoxia-inducible factor 1 alpha - chicken

C/Species: Gallus gallus (chicken)

C/Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 30-Jun-2001



```

Oy 120 CSSL:-----IHNPFGNFS-----LEIGHSPFPIHPCDDELDALTPR 162
Db 175 LKQTMDFGLVLSHGSDIYVSENWVEYIGITMIDTLGQGIWESHCDAHEALSLK 234

Oy 163 PNLSEKKLEAPTE-----RHFSLRKSTLTSRGRTNLKAAATKVLHCSGM 209
Db 235 RELAQKVFQKDEPQGNQSVSTHHRDLFVRLKCTLTLSRGRSINIKASYSKVHIITGHL 289

RESULT 7
A:58520
single-minded gene 2 protein - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 20-Feb-1997 #sequence revision 27-Feb-1997 #text_change 09-Jul-2004
C:Accession: A58520; I58111; I78524; I78523; S61256
R:Osogawa, K.; Okeno, S.; Kato, Y.; Nishimura, Y.; Soeda, E.
A:Title: A 19-kb CGF island associated with single-minded gene 2 in Down syndrome chromosome 21
A:Reference number: A58520; MUID:97061206; PMID:8905236
A:Accession: A58520
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-248 <OS>
A:Cross-references: UNIPROT:Q14190; DDBJ:D85922
N:Chen, H.; Chasac, R.; Rossier, C.; Gos, A.; Antonarakis, S.E.; Kudoh, J.; Yamaki, A.; Saito, Y.
Nature Genet. 10, 9-10, 1995
A:Title: Single-minded and Down syndrome? [letter].
A:Reference number: I58111; MUID:95375794; PMID:7647800
A:Accession: I58111
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-58 <RE>
A:Cross-references: GB:D44444; NID:g1088450; PIDN:BA07906.1; PID:g1088451
A:Accession: I78524
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 117-152 <RE2>
A:Cross-references: GB:D44446; NID:g1088454; PIDN:BA07908.1; PID:g1088455
A:Accession: I78523
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 87-116 <RE3>
A:Cross-references: GB:D44445; NID:g1088452; PIDN:BA07907.1; PID:g1088453
R:Delabar, J.M.; Dahmane, X.Y.Z.
Submitted to the EMBL Data Library, February 1995
A:Description: Down syndrome critical region encodes a gene with homology to Drosophila 87A
A:Reference number: S61256
A:Accession: S61256
A:Status: preliminary
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 87-116 <DEL>
A:Cross-references: EMBL:X84790; NID:g971469; PIDN:CAA59261.1; PID:g971470
A:Accession:
A:Gene: hSIM2
F:1-50/Region: helix-loop-helix #status predicted

Query Match 15 6%; Score 256; DB 2; Length 248;
Best Local Similarity 35.3%; Pred. No. 7, 3e-15;
Matches 73; Conservative 26; Mismatches 78; Indels 30; Gaps 6;

Oy 16 KEKSDAARSRSSQTEVLVQLAHTLPFARGVSAHLDDKASIMRLTTSYLMHRLCAAGK 75
Db 2 KEKSNAAKTRREKNGEYELAKLPLPSAITSQDXXSIINLTSTYLMGRAVFPFG-- 59

Oy 76 RGRATGRLLPBCPGFGRHGTNRGRHGLPV-----GCCQAPGQSVDLCSSSL 124
Db 60 LGDMWGQ--PSRAAGLIDGAKELGSHLLQTLDFVEVVAVDGKIMYISRTASVHLGLS-- 115

Oy 125 IHNPFGNFSLETLGHSIPDFIHPDODELDALTPRPVLSKCKK--EAPTERHFSLRMK 183
Db 116 -----QVELGNSIYIYIIPSDHDEMAVLTAAQPLHHHLQEIYIERSFFLRMK 165

```

Qy 184 STLSRGRTLNKKAATWKVHCSGMR 210  
Db 166 CVLAK--RNAGLTCSGYKVIHCSGYLK 190

## RESULT 8

A29945  
neurogenesis regulatory protein - fruit fly (Drosophila melanogaster) (fragment)  
N:Alternate names: single-minded gene protein  
C:Species: Drosophila melanogaster  
C>Date: 15-Dec-1998 #sequence\_revision 15-Dec-1998 #text\_change 09-Jul-2004  
C:Accession: A29945  
R:Crewe, S.T.; Thomas, J.B.; Goodman, C.S.  
Cell 52, 143-151, 1998  
A>Title: The Drosophila single-minded gene encodes a nuclear protein with sequence simi  
A:Reference number: A29945; MUID:88151023; PMID:3345560  
A:Accession: A29945  
A:Molecule type: mRNA  
A:Residues: 1-655 <CR>  
A:Cross-references: UNIPROT:P05709; GB:M19020; NID:G158464; PID:G158465  
C:Gene: sim  
A:Gene: sim  
C:Keywords: DNA binding; transcription regulation

Query Match 12.3%; Score 202; DB 2; Length 655;  
Best Local Similarity 28.5%; Pred. No. 1.2e-09;  
Matches 73; Conservative 35; Mismatches 78; Indels 70; Gaps 10;

Qy 36 QLAHTLPFAGVSAHMLDKASIMRLTISYLRNRCAGAGKRGRTGRLPEPGGPFHGT 95  
Db 4 ELAKLPLPAATISQLDASVIRLRTISYLRKRVFPDG--LGEAWG---SSPAMQKAT 57  
Qy 96 HRR-GRHGL-----PVGKCOQAPGPOVDLCSSSLIHNPTGTFSLIGHSI 143  
Db 58 IKELGSHLLQTLDFIFVVAADGKIMYISFSAVHLGS-----QVELTNSI 105  
Qy 144 PDIHPDCDELDALTPRPVLSKKL-----EAPL-----ERH 177  
Db 106 FEYIHNVDDEMNAILSLHPINOHPLAQHTTPIGSPGVQHPSPAYHDHSGHTIELEKT 165  
Qy 178 FSLMKSTLSRGRTLNKKAATWKVHCSGMRAYKPPAQGSPRSPRPPQLQCLVLC 237  
Db 166 FFLMKCVLAK--RNAGLTISGFVTHCSGYLKARIYPRDGSGS-----LIONLGLVA 218  
Qy 238 -----EALPOLPFH 246  
Db 219 VGHSLPSSAITEIKLH 234

## RESULT 9

T21943  
hypochemical protein F38A6.3a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T21943; T26899  
R:Mortimore, B.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19491  
A:Accession: T21943  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-823 <W1>  
A:Cross-references: UNIPROT:O45486; EMBL:Z92833; PIDN:CAB07380.1; GSPDB:GN00023; CESP:F3  
A:Experimental source: clone F38A6  
R:Ainscough, R.  
submitted to the EMBL Data Library, June 1998  
A:Reference number: Z20282  
A:Accession: T26899  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-823 <W1>  
A:Cross-references: EMBL:AL023842; PIDN:CA119520.1; GSPDB:GN00023; CESP:F38A6.3a

A:Experimental source: clone Y44A6D  
C:Gene: CESP:F38A6.3a  
A:Map position: 5  
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 711/1; 768/3

Query Match 10.4%; Score 169.5; DB 2; Length 823;  
Best Local Similarity 26.3%; Pred. No. 1.1e-06;  
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

Qy 9 RSTTELRKESRDAASRSQETEVLYOAHLPFA-RGVSAHLDKASIMRLTISYLRN 67  
Db 110 KRMEKRRRTSRDAADRSKESDIPDDLMCVIYEBGVTHLDIALLRVAATTCRLR 169  
Qy 68 RLCA---AGGKRGATGRLLPEPGGFRHGTNRGRHGLPVGKCOQAPGPOVDLCSSSL 124  
Db 170 KTAGNVLENLNDNEITNEWTED-----TIAECLDG-FVMIVDSDSL 212  
Qy 125 IHNPTGTFSL---ELIGHSTFDIHPDCDELDALTPRPVLSKKLEAP-----T 174  
Db 213 YVTESVAMYGLQTLGRALRDFLHPSDYDEF-----DKQSKMLHKRPGEDTDTT 264  
Qy 175 ERHFSLRMKSTLSRGRTLNKKAATWK---VLH-----CSGMRAYKPPAQGSPAG 222  
Db 265 GIMVLRMKTVISPRGRCLNLSALYKSVFLVHRSVSTGHSFPG--GITTPAG 318

## RESULT 10

T21944  
hypochemical protein F38A6.3b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T21944; T26900  
R:Mortimore, B.  
submitted to the EMBL Data Library, March 1997  
A:Reference number: Z19491  
A:Accession: T21944  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-825 <W1>  
A:Cross-references: UNIPROT:Q9TVN0; EMBL:Z92833; PIDN:CAB07381.1; GSPDB:GN00023; CESP:F3  
A:Experimental source: clone F38A6  
R:Ainscough, R.  
submitted to the EMBL Data Library, June 1998  
A:Reference number: Z20282  
A:Accession: T26900  
A:Molecule type: DNA  
A:Residues: 1-825 <W1>  
A:Cross-references: EMBL:AL023842; PIDN:CA119521.1; GSPDB:GN00023; CESP:F38A6.3b  
A:Experimental source: clone Y44A6D  
A:Gene: CESP:F38A6.3b  
A:Map position: 5  
A:Introns: 19/2; 75/1; 111/2; 175/1; 226/3; 331/3; 518/1; 627/2; 713/1; 770/3

Query Match 10.4%; Score 169.5; DB 2; Length 825;  
Best Local Similarity 26.3%; Pred. No. 1.1e-06;  
Matches 62; Conservative 35; Mismatches 90; Indels 49; Gaps 10;

Qy 9 RSTTELRKESRDAASRSQETEVLYOAHLPFA-RGVSAHLDKASIMRLTISYLRN 67  
Db 110 KRMEKRRRTSRDAADRSKESDIPDDLMCVIYEBGVTHLDIALLRVAATTCRLR 169  
Qy 68 RLCA---AGGKRGATGRLLPEPGGFRHGTNRGRHGLPVGKCOQAPGPOVDLCSSSL 124  
Db 170 KTAGNVLENLNDNEITNEWTED-----TIAECLDG-FVMIVDSDSL 212  
Qy 125 IHNPTGTFSL---ELIGHSTFDIHPDCDELDALTPRPVLSKKLEAP-----T 174  
Db 213 YVTESVAMYGLQTLGRALRDFLHPSDYDEF-----DKQSKMLHKRPGEDTDTT 264  
Qy 175 ERHFSLRMKSTLSRGRTLNKKAATWK---VLH-----CSGMRAYKPPAQGSPAG 222

Db 265 GINNVLRKVTISPGRCCLNKLKSLYKSVSFLVHSKYSTGCHVSPMQ--GITTPAG 318

RESULT 11

JC7635

aryl hydrocarbon receptor nuclear translocator 1 - chicken

N/Alternate names: hypoxia-inducible factor 1 beta

C/Species: Gallus gallus (chicken)

C/Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004

C/Accession: JC7635

R/Cation: T.; Mentola, M.A.; Smith, S.M.; Born, J.; Walker, M.K.

Biochem. Biophys. Res. Commun. 282, 602-607, 2001

A/Title: Hypoxia regulates avian cardiac Arnt and HIF-1alpha mRNA expression.

A/Reference number: JC7635; PMID:21294777; PMID:11401503

A/Contents: Embryo

A/Accession: JC7635

A/Molecule type: mRNA

A/Residues: 1-805 <CAT>

A/Cross-references: UNIPROT:Q98SN3

C/Comment: This protein, known as hypoxia-inducible factor 1 beta, a member of the basic pones to hormonal and environmental stimuli. It is also a common dimer partner for trar

C/Genetics:

A/Gene: arntl

C/Keywords: transcription factor

Query Match 8.7%; Score 142; DB 2; Length 805;  
 Best Local Similarity 20.5%; Pred. No. 0.00027;  
 Matches 61; Conservative 43; Mismatches 95; Indels 98; Gaps 8;

Qy 14 LRKESRDAAARRSQETEVLYOLAHTLPFARGVSAHLDAKASIMRLTISYLRMRCLCAAG 73  
 Db 89 LARENHSEIERRRRNKMTAYITELSDVPTCSALARKDKLTILRMVASHM----- 139

Qy 74 GKRGRATGRLLPBGPGFRRGTHRGHGLPVKCCQAQPGQSVDLCSLSLI---HNPT 129  
 Db 140 -KSLRGTGNTSTDGKYPSPFLTDDELKHLI---LEADGFLFIVSCETGRVYVSDSVT 194

Qy 130 PGTNF-SLELIGHISIPDFIHPDCOELODALTPRPN-----LSKKLEAPTRRH 178  
 Db 195 PVLNQPSWFGSTLVQVHPDDVDKLRQLSTSESNALTBGTXPWCLSNDPAPRPS-- 252

Qy 179 SLRKSTLTLSRKRTLLKATWK----- 201  
 Db 253 -----ASKGRILDLKTGTVKKEGQOSMRGMSRRSFCRMRCGNSVDVAVSVRLS 304

Qy 202 -----VLKSGHMAKYPKPAQSPSPGSP-RSEPPQLCVLI 236  
 Db 305 FMNRRCNGLGATYDGEHPHYVVAHCTGYIKAMPAGVSLPDDPDAGGSKFCIVAI 361

RESULT 12

JC7633

aryl hydrocarbon nuclear translocator ARNT2-like factor, ARNT2X - zebra fish

C/Species: Brachydanio rerio (zebra fish)

C/Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004

C/Accession: JC7633

R/Hau, H.U.; Wang, W.D.; Hu, C.H.

Biochem. Biophys. Res. Commun. 282, 487-492, 2001

A/Title: Ectopic expression of negative ARNT2 factor disrupts fish development.

A/Reference number: JC7633; PMID:21294759; PMID:11401485

A/Contents: Heart

A/Accession: JC7633

A/Molecule type: mRNA

A/Residues: 1-392 <HSU>

A/Cross-references: UNIPROT:Q98SK3; GB:AY007992

C/Comment: This factor, a basic helix-loop-helix PAS (bHLH-PAS) factor, heterodimerizes

C/Genetics:

A/Gene: arnt2X

Query Match 8.5%; Score 138.5; DB 2; Length 392;  
 Best Local Similarity 20.8%; Pred. No. 0.00023;  
 Matches 55; Conservative 48; Mismatches 103; Indels 59; Gaps 8;

Qy 16 KEKSRDAAARRSQETEVLYOLAHTLPFARGVSAHLDAKASIMRLTISYLRMRCLCAAGK 75  
 Db 47 RENNSEIERRRRNKMTQYTETELSDMPTCSALARKDKLTILRMVASHM-----K 96

Qy 76 RGRATGRLLPBGPGFRRGTHRGHGLPVKCCQAQPGQSVDLCSLSLI---HNPTPG 131  
 Db 97 SMRGTGNTSTDGAKYKPSFLTEDELKHLI---LEADGFLFVVAETGRVYVSDSVTPV 152

Qy 132 TNF-SLELIGHISIPDFIHPDCOELODALTPRPVLSKKKL----- 170  
 Db 153 LNHQSWFGSTLVQVHPDDVDKLRQLSTSESNALTBGTXPWCLSNDPAPRPS-- 212

Qy 171 -----EAPTRHPSLRKSTLTSSRGT-----LNLKVAATWKYLHSGHMRAY 212  
 Db 213 MGRSRSPICMRGCSALD-HISLNRSSMRKRYRNLGSPKGEAQYSVVHCTGYIKAW 271

Qy 213 KPPAQITSP-AGSPREPPLOCLVLI 236  
 Db 272 PPAGMTIPDEDTEAGQTSKYCLVAI 296

RESULT 13

A56241

aryl hydrocarbon receptor nuclear translocator protein - mouse

C/Species: Mus musculus (house mouse)

C/Date: 03-Oct-1995 #sequence\_revision 03-Oct-1995 #text\_change 09-Jul-2004

C/Accession: A56241

R/Reisz-Porzsasz, S.; Probat, M.R.; Fukunaga, B.N.; Hankinson, O.

Mol. Cell. Biol. 14, 6075-6086, 1994

A/Title: Identification of functional domains of the aryl hydrocarbon receptor nuclear

A/Reference number: A56241; PMID:94344118; PMID:8065341

A/Accession: A56241

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-791 <REI>

A/Cross-references: UNIPROT:P53762; GB:U10325; NID:G555687; PID:AAA5671.1; PID:G555688

A/Note: authors translated the codon TTT for residue 375 as Ser, TGT for residue 380 as

A/Note: authors failed to translate CAG for residue 507 as Gln

C/Genetics:

A/Gene: ARNT

Query Match 8.5%; Score 138.5; DB 2; Length 791;  
 Best Local Similarity 21.8%; Pred. No. 0.00053;  
 Matches 58; Conservative 50; Mismatches 101; Indels 57; Gaps 9;

Qy 14 LRKESRDAAARRSQETEVLYOLAHTLPFARGVSAHLDAKASIMRLTISYLRMRCLCAAG 73  
 Db 89 LARENHSEIERRRRNKMTAYITELSDVPTCSALARKDKLTILRMVASHM----- 139

Qy 74 GKRGRATGRLLPBGPGFRRGTHRGHGLPVKCCQAQPGQSVDLCSLSLI---HNPT 129  
 Db 140 -KSLRGTGNTSTDGKYPSPFLTDDELKHLI---LEADGFLFIVSCETGRVYVSDSVT 194

Qy 130 PGTNF-SLELIGHISIPDFIHPDCOELODALTPRPN-LSKKLEAPTR----- 174  
 Db 195 PVLNQPSWFGSTLVQVHPDDVDKLRQLSTSESNALTBGTXPWCLSNDPAPRPS-- 254

Qy 175 -----EHPESLRKSTLTSSRGT-----RGRTLNLKAT-----WKYHLCSGHMRA 211  
 Db 255 MCGSRSPICMRGCSALD-HISLNRSSMRKRYRNLGSPKGEAQYSVVHCTGYIKA 314

Qy 212 YKPPAQITSP-AGSPREPPLOCLVLI 236  
 Db 315 WPPAGVSLPDDPDAGGSKFCIVAI 340

RESULT 14

A55448

Ah receptor nuclear translocator - mouse

C/Species: Mus musculus (house mouse)

C/Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 09-Jul-2004

C/Accession: A55448

R, Li, H.; Dong, L.; Whitlock Jr., J.P.  
 J. Biol. Chem. 269, 28098-28105, 1994  
 A>Title: Transcriptional activation function of the mouse Ah receptor nuclear translocator  
 A;Reference number: A55448; MUID:95050586; PMID:7961746  
 A;Accession: A55448  
 A>Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-776 <LIA>  
 A;Cross-references: UNIPROT:Q921F3; GB:U14333

Query Match 8.3%; Score 136.5; DB 2; Length 776;  
 Best Local Similarity 21.2%; Pred. No. 0.00077;  
 Matches 58; Conservative 50; Mismatches 109; Indels 57; Gaps 9;

QY 6 QRVASNTLRKRSRDARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLR 65  
 DB QMCDKERPARENNSEIERRRRNKMTAVITELSDMVPFCSALARKPKDLTLRMAVSHM- 124  
 QY 66 MHRICAAAGKRGATGRLLPBGPGGFRHGTNRGRHGLPVGKCOQAPGQSVLDCSSLI 125  
 DB 125 -----KSLRGTGNTSDGYKPSFLTDQELKHLI-----LEADGFLFVSCETGRV 171  
 QY 126 ----HNPTPGTNF-SLELIGHSTFDLHPCDQELQDALTPRPV-LSKKKLEAPT----- 174  
 DB 172 VYVSDSVTPVLNQPSQSWFGSTLYDVHPDQVDRKLRQLSTSENALTGRLDLKGTGKK 231  
 QY 175 -----ERHFSLRMKSTLTS-----RGRTLNLKAAAT-----WKVL 203  
 DB 232 EGGQSSMRMCMGSRSPFCMRRCCTSSVDVPVSMRSLFLNRRCNGLGSVKEGPHFVVV 291  
 QY 204 HCSGMRAYKPPAQTSPAGSPRS-EPPLOCLVLI 236  
 DB 292 HCTGYIKAMPAGVSLPDDPDEAGQSGKFCIVAI 325

## RESULT 15

I59550  
 aryl hydrocarbon receptor nuclear translocator Arnt [imported] - human  
 C;Species: Homo sapiens (man)  
 C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
 C;Accession: I59550  
 R;Hoffman, E.C.; Reyes, H.; Chu, F.F.; Sander, F.; Conley, L.H.; Brooks, B.A.; Hankinson  
 Science 252, 954-958, 1991  
 A>Title: Cloning of a factor required for activity of the Ah (dioxin) receptor.  
 A;Reference number: I59550; MUID:91240280; PMID:1852076  
 A;Accession: I59550  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-789 <RBS>  
 A;Cross-references: UNIPROT:P27540; GB:M69238; NID:G179003; PIDN:AAA51777.1; PID:G179004  
 C;Genetics:  
 A;Gene: GDB:ARNT  
 A;Cross-references: GDB:119701; OMIM:126110  
 A;Map position: 1q21-1q21

Query Match 8.3%; Score 136.5; DB 2; Length 789;  
 Best Local Similarity 21.8%; Pred. No. 0.00079;  
 Matches 58; Conservative 50; Mismatches 101; Indels 57; Gaps 9;

QY 14 LRKESDAARSRRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMHRICAAAG 73  
 DB 89 LARENHSEIERRRRNKMTAVITELSDMVPFCSALARKPKDLTLRMAVSHM----- 139  
 QY 74 GKRGATGRLLPBGPGGFRHGTNRGRHGLPVGKCOQAPGQSVLDCSSLI-----HNPT 129  
 DB 140 -KSLRGTGNTSDGYKPSFLTDQELKHLI-----LEADGFLFVSCETGRVYVSDSVT 194  
 QY 130 PGTNF-SLELIGHSTFDLHPCDQELQDALTPRPV-LSKKKLEAPT----- 174  
 DB 195 PVLNQPSQSWFGSTLYDVHPDQVDRKLRQLSTSENALTGRLDLKGTGKVEGQSSMR 254  
 QY 175 -----ERHFSLRMKSTLTS-----RGRTLNLKAAAT-----WKVLHCSGMR 211

DB 255 MCKGSRSPFCMRRCSSSVDPVSVNRLSFVNRRCNGLGSYKDGEPHFVVAHCTGYIKA 314  
 QY 212 YKPPAQTSPAGSPRS-EPPLOCLVLI 236  
 DB 315 WPPAGVSLPDDPDEAGQSGKFCIVAI 340

Search completed: June 15, 2005, 16:00:54  
 Job time : 42 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 15:37:58 ; Search time 173 Seconds  
(without alignments)  
908.719 Million cell updates/sec

Title: US-09-896-791B-3  
Perfect score: 1636  
Sequence: 1 MALGLQVRNTELRKEKSR.....TESSLPSWVLMALNRKNCPCG 307

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_03.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1636	100.0	307	Q8VHR1	Q8vhr1 mus musculus
2	855.5	52.3	630	Q9GX54	Q9gx54 mus musculus
3	855.5	52.3	662	Q9Z215	Q9z215 mus musculus
4	835	51.0	632	Q8WXA1	Q8wx1 homo sapien
5	835	51.0	669	Q6KX72	Q6kx72 homo sapien
6	831	50.8	662	Q96K34	Q96k34 homo sapien
7	820.5	50.2	662	Q9JH52	Q9jh52 ratius norv
8	802	49.0	667	Q9Y2N7	Q9y2n7 homo sapien
9	573	35.0	648	Q9HAI2	Q9hai2 homo sapien
10	518.5	31.7	199	Q9HAMS	Q9ham5 homo sapien
11	503	30.7	811	HIFA_CHICK	O9y1b5 gallus galli
12	499	30.5	823	HIFA_BOVIN	O9xtab5 bos taurus
13	499	30.5	823	O6IV47	O6iv47 bos mutus g
14	497	30.4	825	HIFA_RAT	O35800 ratius norv
15	495	30.3	489	Q6EMT3	O6emt3 xenopus lae
16	495	30.3	826	HIFA_HUMAN	O16665 homo sapien
17	495	30.3	836	HIFA_MOUSE	O61221 mus musculus
18	493	30.1	802	O6P154	O6p154 xenopus lae
19	492.5	30.1	643	O6STN6	O6stn6 ctenopharyn
20	492	30.1	819	Q7YSE5	O7yse5 cycloclagus
21	490.5	30.0	777	Q6EH14	Q6eh14 brachydanio
22	489.5	29.9	774	O6STN7	O6stn7 ctenopharyn
23	488	29.8	874	PAS1_MOUSE	P97481 mus musculus
24	488	29.8	874	O6PBU2	O6pbu2 mus musculus
25	484	29.6	821	O64FS4	O64fs4 spermophilu
26	483.5	29.6	533	O6P133	O6p133 brachydanio
27	482	29.5	874	PAS1_RAT	O9jha1 ratius norv
28	481.5	29.4	626	O6EGR9	O6egr9 brachydanio
29	480	29.3	786	O6SLI1	O6sli1 canis famil
30	479	29.3	835	O6P6W2	O6p6w2 ctenopharyn
31	478	29.2	824	O6H8T3	O6h8t3 spalax juda

32	474.5	29.0	873	Q8QGM4	Q8qgm4 fundulus he
33	471	28.8	862	O6GL61	O6gl61 xenopus tro
34	471	28.8	867	O9W7C6	O9w7c6 gallus galli
35	471	28.8	870	O9PTB3	O9ptb3 coturnix co
36	468	28.6	870	O9XTA4	O9xta4 bos taurus
37	467	28.5	870	PAS1_HUMAN	Q99814 homo sapien
38	464	28.4	862	O6GP97	O6gp97 xenopus lae
39	462	28.2	859	O6GQ12	O6gq12 xenopus lae
40	450	27.5	766	HIFA_ONCMY	O988w2 oncorhynchu
41	426	26.0	237	O9H7Z9	O9h7z9 homo sapien
42	389	23.8	189	O6RYC8	O6ryc8 ovis aries
43	376	23.0	189	O6RYC9	O6ryc9 sus scrofa
44	360	22.0	65	O8R4D6	O8r4d6 mus musculus
45	323.5	19.8	571	O7T2E4	O7t2e4 brachydanio

ALIGNMENTS

RESULT 1	ID	Q8VHR1	PRELIMINARY;	PRT;	307 AA.
AC	Q8VHR1	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)				
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)				
DE	Inhibitory PAS domain protein.				
GN	Name=Hif3a, Synonyms=Idpas;				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.				
OX	NCBI_TaxID=10090;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=C57BL/6J;				
RX	PubMed=11734856; DOI=10.1038/35107085;				
RA	Makino Y., Cao R., Svensson K., Bertilsson G., Asman M., Tanaka H.,				
RA	Cao Y., Berkenstam A., Poellinger L.;				
RT	"Inhibitory PAS domain protein is a negative regulator of hypoxia-				
RT	inducible gene expression."				
RL	Nature 414:550-554(2001).				
CC	-1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.				
DR	EMBL; AF416641; ALJ39015.1; -.				
DR	MED; MGI:1859778; Hif3a.				
DR	GO; GO:0005634; C:nucleus; IC.				
DR	GO; GO:0003700; P:transcription factor activity; IPI.				
DR	GO; GO:0001666; P:response to hypoxia; IDA.				
DR	PFam; PF00010; HLH; 1.				
DR	SMART; SM00353; HLH; 1.				
SQ	SEQUENCE 307 AA; 33990 MW; D89D3B25C9C3967 CRC64;				
Query Match	100.0%;	Score 1636;	DB 2;	Length 307;	
Best Local Similarity	100.0%;	Pred. No. 1.2e-127;			
Matches 307;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	MALGLQVRNTELRKEKSRDAARSRSSQETVYLQALHTLPARGVSAHLDKASIRLT	60		
DB	1	MALGLQVRNTELRKEKSRDAARSRSSQETVYLQALHTLPARGVSAHLDKASIRLT	60		
QY	61	ISYLRMRRLCAAGKGRATGRLLPEGPGFRRGTHRRGHGLPVGCKQQAPOGVSVDLC	120		
DB	61	ISYLRMRRLCAAGKGRATGRLLPEGPGFRRGTHRRGHGLPVGCKQQAPOGVSVDLC	120		
QY	121	SSSLINHPPTGNTFSLDELIGHSIFDFIHPCDQBELDALTFRPNLSKKKLEAPTERFSL	180		
DB	121	SSSLINHPPTGNTFSLDELIGHSIFDFIHPCDQBELDALTFRPNLSKKKLEAPTERFSL	180		
QY	181	RMSTLTLSRRTNLKATKVLHSGHMRAYYPAPQTSFAGSPRSPPLOCTVLICEAI	240		
DB	181	RMSTLTLSRRTNLKATKVLHSGHMRAYYPAPQTSFAGSPRSPPLOCTVLICEAI	240		
QY	241	POLPFHDGATLGPQEKPTISTLFTPLMKALLCLVKKMPVOVLQKGTESSLPSWVLMAL	300		

Db 241 POLPFHGGATLGLPQEKPISTLPTPLMKALLCLVKEWPVQVLGGKTESSLSPLVLMAL 300  
 QY 301 NRKNCPG 307  
 Db 301 NRKNCPG 307

## RESULT 2

090X54 ID 090X54 PRELIMINARY; PRT; 630 AA.

AC 090X54; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Hypoxia-inducible factor 3 alpha (Fragment).  
 GN Name=Hif3a;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]

RP SEQUENCE FROM N.A.

RA Gu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
 RT "Cloning and Characterization of a Third Hypoxia Inducible Factor,  
 RT HIF3-alpha."  
 RL J. Biol. Chem. 0:0-0(1999).  
 RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=99054547; PubMed=9840812;  
 RA Gu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
 RT "Molecular characterization and chromosomal localization of a third  
 RT alpha-class hypoxia inducible factor subunit, HIF3alpha."  
 RL Gene Expr. 7:205-213(1998).  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

DR EMBL; AF079153; AAF21782.1; -  
 DR EMBL; AF079140; AAF21782.1; JOINED.  
 DR EMBL; AF079143; AAF21782.1; JOINED.  
 DR EMBL; AF079143; AAF21782.1; JOINED.  
 DR EMBL; AF079145; AAF21782.1; JOINED.  
 DR EMBL; AF079147; AAF21782.1; JOINED.  
 DR EMBL; AF079149; AAF21782.1; JOINED.  
 DR EMBL; AF079151; AAF21782.1; JOINED.  
 DR EMBL; AF079152; AAF21782.1; JOINED.  
 DR EMBL; AF079150; AAF21782.1; JOINED.  
 DR EMBL; AF079148; AAF21782.1; JOINED.  
 DR EMBL; AF079146; AAF21782.1; JOINED.  
 DR EMBL; AF079144; AAF21782.1; JOINED.  
 DR EMBL; AF079142; AAF21782.1; JOINED.  
 DR HSSP; Q99814; 1P97.  
 DR MGD; MGI:1859778; Hif3a.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003700; F:transcription factor activity; IPI.  
 DR GO; GO:0001666; P:response to hypoxia; IDA.  
 DR GO; GO:0006366; P:transcription from Pol II promoter; IPI.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR InterPro; IPR001067; NUC\_translocat.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF000010; HLH; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR PRINTS; PR00785; NCTRNSLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PSS0112; PAS; 2.  
 FT NON\_TER 630  
 SQ SEQUENCE 630 AA; 69623 MW; 828EB2CB4B6D45B6 CRC64;

Query March 52.3%; Score 855.5; DB 2; Length 630;  
 Best Local Similarity 68.4%; Pred. No. 1,8e-62;  
 Matches 182; Conservative 5; Mismatches 24; Indels 55; Gaps 4;

QY 9 RSTTELRKESKRDPAASRSQETEVLYOLAHITLPPARGVSAHLDKASIMRLTISYLMR 68  
 Db 7 RSTTELRKESKRDPAASRSQETEVLYOLAHITLPPARGVSAHLDKASIMRLTISYLMR 66  
 QY 69 LCAAGGGRGATGRLRLEGGCGFRHGRGRGLGVKCGQQAPEGQSVDLG----- 120  
 Db 67 LCAAGGWN-----QVEKGGEPLDCTLKALGCF 94  
 QY 121 -----SSSLHNPTPEGNFS-DELIGHSIFDFIHPDQDELQDALTPRNLSSKKL 170  
 Db 95 VMLTLAGDMAYLSENVSKLIGLSQELIGHSIFDFIHPDQDELQDALTPRNLSSKKL 154  
 QY 171 EATTEHFSLRKSTLTSGRTINLKAATWKVLHSGHNRAYRPPAQTSFAGSPREPP 230  
 Db 155 EATTEHFSLRKSTLTSGRTINLKAATWKVLHSGHNRAYRPPAQTSFAGSPREPP 214  
 QY 231 QCLVLCEATP-----QLPFHGGATL 251  
 Db 215 QCLVLCEATLPPASLEPPLGRCAPL 240

## RESULT 3

092Z15 ID 092Z15 PRELIMINARY; PRT; 662 AA.

AC 092Z15; 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Hypoxia inducible factor three alpha.  
 GN Name=Hif3a;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99054547; PubMed=9840812;  
 RA Gu Y.-Z., Moran S.M., Hogenesch J.B., Wartman L., Bradfield C.A.;  
 RT "Molecular characterization and chromosomal localization of a third  
 RT alpha-class hypoxia inducible factor subunit, HIF3alpha."  
 RL Gene Expr. 7:205-213(1998).  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

DR EMBL; AF060194; AAC72734.1; -  
 DR HSSP; Q99814; 1P97.  
 DR MGD; MGI:1859778; Hif3a.  
 DR GO; GO:0005634; C:nucleus; IC.  
 DR GO; GO:0003700; F:transcription factor activity; IPI.  
 DR GO; GO:0001666; P:response to hypoxia; IDA.  
 DR GO; GO:0006366; P:transcription from Pol II promoter; IPI.  
 DR InterPro; IPR001092; HLH\_basic.  
 DR InterPro; IPR001067; NUC\_translocat.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF000010; HLH; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR PRINTS; PR00785; NCTRNSLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PSS0112; PAS; 2.  
 SQ SEQUENCE 662 AA; 73012 MW; 58740A1B6993D3B5 CRC64;

Query March 52.3%; Score 855.5; DB 2; Length 662;  
 Best Local Similarity 68.4%; Pred. No. 2e-62;  
 Matches 182; Conservative 5; Mismatches 24; Indels 55; Gaps 4;

QY 9 RSTTELRKESKRDPAASRSQETEVLYOLAHITLPPARGVSAHLDKASIMRLTISYLMR 68  
 Db 7 RSTTELRKESKRDPAASRSQETEVLYOLAHITLPPARGVSAHLDKASIMRLTISYLMR 66  
 QY 69 LCAAGGGRGATGRLRLEGGCGFRHGRGRGLGVKCGQQAPEGQSVDLG----- 120  
 Db 67 LCAAGGWN-----QVEKGGEPLDCTLKALGCF 94



Qy 121 -----SSSLINHPPTGTFNS-LELIGHSIFDPIHPCQDELODALTPRPNISSKKKL 170  
 Db 95 VVWVLTAEQDMVYLSNENSKHGLSGLDELIGHSIFPIHPCQDELODALTPRPNISSKKKL 154  
 Qy 171 EAPFERHSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 230  
 Db 155 EAPFERHSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 214  
 Qy 231 QCLVLTCEAIP-----OLPFDGATL 251  
 Db 215 QCLVLTCEAIPHPASLEPPLGRGATL 240

RESULT 4

08WKX1 PRELIMINARY; PRT; 632 AA.  
 AC 08WKX1.  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Inhibitory PAS domain protein.  
 GN Name:IPAS;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cheng J.Q.;  
 RL Submitted (DDBJ-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 DR EMBL; AF434942; AAL69947.1; -.  
 DR HSSP; O16655; 1IQB  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0004871; F:signal transducer activity; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00989; PAS; 1.  
 DR PRINTS; PR00785; NCTRNLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PSS0112; PAS; 2.  
 SQ SEQUENCE 632 AA; 68963 MW; 9665BDAF3998F8EF CRC64;

Query Match 51.0%; Score 835; DB 2; Length 632;  
 Best Local Similarity 68.6%; Pred. No. 9.3e-61;  
 Matches 179; Conservative 11; Mismatches 29; Indels 42; Gaps 5;  
 Qy 1 MALGLQVRNSTELRKEKSRDAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 Db 1 MALGLQVRNSTELRKEKSRDAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 Qy 61 ISYLRMRRLCAAGKRRGRATGRLLPEBPGGFRHGTNRHGRGLPVGKQ----- 109  
 Db 61 ISYLRMRRLCAAG-----EKNVGAAGEPLDACYLKALGFWVL 100  
 Qy 110 QAPPGQVDLCSSSLINHPPTGTFNS-LELIGHSIFDPIHPCQDELODALTPRPNISSKK 168  
 Db 101 TAEGDMVY-----LSENVSKHGLSGLDELIGHSIFPIHPCQDELODALTPQQTLSRR 154  
 Qy 169 KLEAPTERHSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 228  
 Db 155 KVEAPTERCFSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 214  
 Qy 229 PLOCLVLTCEAIPOLPFHDGA 249  
 Db 215 PLOCLVLTCEAIP-----HPGS 231

RESULT 5

066K72 PRELIMINARY; PRT; 669 AA.  
 AC 066K72.  
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE HIF3A protein.  
 GN Name-HIF3A;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NC NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RX PubMed:12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Ditchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muliyil S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalek U., Smalhe U., Smalhe D.E., Schenck A., Schein J.E.,  
 RA Jones S.J., Maitra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Pancreas;  
 RA Director MGC Project;  
 RL Submitted (Aug-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 DR EMBL; BC080551; AAB80551.1; -.  
 DR InterPro; IPR001092; HLH basic.  
 DR InterPro; IPR001067; NUC\_translocat.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF00010; HLH; 1.  
 DR PRINTS; PR00785; NCTRNLOCATR.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PSS0112; PAS; 2.  
 SQ SEQUENCE 669 AA; 72460 MW; 7EF7362A1691AED CRC64;

Query Match 51.0%; Score 835; DB 2; Length 669;  
 Best Local Similarity 68.6%; Pred. No. 1e-60;  
 Matches 179; Conservative 11; Mismatches 29; Indels 42; Gaps 5;  
 Qy 1 MALGLQVRNSTELRKEKSRDAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 Db 1 MALGLQVRNSTELRKEKSRDAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 Qy 61 ISYLRMRRLCAAGKRRGRATGRLLPEBPGGFRHGTNRHGRGLPVGKQ----- 109  
 Db 61 ISYLRMRRLCAAG-----EKNVGAAGEPLDACYLKALGFWVL 100  
 Qy 110 QAPPGQVDLCSSSLINHPPTGTFNS-LELIGHSIFDPIHPCQDELODALTPRPNISSKK 168  
 Db 101 TAEGDMVY-----LSENVSKHGLSGLDELIGHSIFPIHPCQDELODALTPQQTLSRR 154  
 Qy 169 KLEAPTERHSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 228  
 Db 155 KVEAPTERCFSLRMKSTLTSTRGRTLNKAAATWVYLHCSGMRAYKPPAQTSPAGSPSEPP 214



```

Db 7 RSSTELKREKSRDARSRSOETVLYQLAHTLPFAKVSADLKASIMRLTISYLMNR 66
Qy 69 LCAAGKRGKRGATGRLPEPGGFRHGTBRGRHGLPVGKCQ-----QAPGPOSV 117
Db 67 LCAAG-----EMNQVKGGEPRIDACTYKALBGFVWVLTBGMAY 106
Qy 118 DLCSSTLIHNPPTGTFNS-LELIGHSIFDFIHPDQBELDALTTPRNLSKKLEAPTER 176
Db 107 -----LSENVSKHLGSQLLELIGHSIFDFIHPDQBELDALTTPRPLSKKSEBAATGR 160
Qy 177 HFSLRMSTLTSRGTINLKAATWKVLCGSHMARAYKPPAOTSPAGSPRSEPPLOCLVLI 236
Db 161 HFSLRMSTLTSRGTINLKAATWKVLCGSHMARAYKPPAOTSPAGSPRSEPPLOCLVLI 220
Qy 237 CEALP-----OLPFPDGT 251
Db 221 CEALPHPASLEPPLGRGAF 240

```

RESULT 8

```

ID 09Y2N7 PRELIMINARY; PRT; 667 AA.
AC 09Y2N7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Putative homolog of hypoxia inducible factor three alpha (Hypoxia-
inducible factor-3 alpha).
GN Name=HIF-3A;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney; PubMed=11573933; DOI=10.1006/dbrc.2001.5659;
RX MEDLINE=21458277;
RA Hara S., Hamada J., Kobayashi C., Kondo Y., Imura N.;
RT "Expression and characterization of hypoxia-inducible factor (HIF) -
3alpha in human kidney: suppression of HIF-mediated gene expression by
HIF-3alpha."
RL Biochem. Biophys. Res. Commun. 287:808-813(2001).
CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
DR EMBL; AC007123; AAD22668.1; -
DR EMBL; AB054067; BAB69689.1; -
DR PIR; JC7771; JC7771.
DR HSSP; Q16665; ILQ8.
DR Genew; HGNC:15825; HIF3A.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR001092; HLH basic.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR Pfam; PF00010; HLH; 1.
DR Pfam; PF00989; PAS; 1.
DR PRINTS; PR00785; NCTRNSLOCATR.
DR SMART; SM00353; HLH; 1.

```

```

DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS00112; PAS; 2.
SQ SEQUENCE 667 AA; 72404 MW; 67B8794FF9DCCF48 CRC64;

```

```

Query Match 49.0%; Score 802; DB 2; Length 667;
Best Local Similarity 68.0%; Pred. No. 5.5e-58;
Matches 172; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

```

```

Qy 9 RSSTELKREKSRDARSRSOETVLYQLAHTLPFAKVSADLKASIMRLTISYLMNR 68
Db 7 RSSTELKREKSRDARSRSOETVLYQLAHTLPFAKVSADLKASIMRLTISYLMNR 66
Qy 69 LCAAGKRGKRGATGRLPEPGGFRHGTBRGRHGLPVGKCQ-----QAPGPOSV 117
Db 67 LCAAG-----EMNQVKGGEPRIDACTYKALBGFVWVLTBGMAY 106
Qy 118 DLCSSTLIHNPPTGTFNS-LELIGHSIFDFIHPDQBELDALTTPRNLSKKLEAPTER 176
Db 107 -----LSENVSKHLGSQLLELIGHSIFDFIHPDQBELDALTTPQOTLSRKVEAPTER 160
Qy 177 HFSLRMSTLTSRGTINLKAATWKVLCGSHMARAYKPPAOTSPAGSPRSEPPLOCLVLI 236
Db 161 CFSLRMSTLTSRGTINLKAATWKVLCGSHMARAYKPPAOTSPAGSPRSEPPLOCLVLI 220
Qy 237 CEALPQLPFDGA 249
Db 221 CEALP-----HPGS 229

```

RESULT 9

```

ID 09HA12 PRELIMINARY; PRT; 648 AA.
AC 09HA12;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FJ11591.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Whole embryo; PubMed=14702039; DOI=10.1038/ng1285;
RX Medline=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,
RA Sekine M., Oobayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahari K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hosiiri T., Kaku Y., Kodaaira H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,
RA Niimomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Horita T., Kusano J.,
RA Kanehori K., Takahashi Fujii A., Hara H., Tanase T., Nomura Y.,
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,
RA Mitsuhashi K., Yuuki H., Oshino K., Sasaki N., Aotsuka S.,
RA Yoshikawa J., Matsunawa H., Ichihara T., Shiohata N., Sano S.,
RA Moriya S., Momiya A., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakeda H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujimori T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,
RA Kawabata A., Hikiji T., Kobayashi N., Inagaki H., Ikema Y., Okamoto S.,
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu T.,
RA Mitsuhashi Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,

```

RA Nakai K., Yada T., Nakamura Y., Ohara O., Isozaki T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RA Nat. Genet. 36:40-45(2004).  
RA EMBL: AK021653; BAB13865.1; -.  
RA HSPB: Q16665; IL08.  
DR GC: GO:0004871; P: signal transducer activity; IEA.  
DR GC: GO:0006355; P: regulation of transcription, DNA-dependent; IEA.  
DR GO: GO:0007165; P: signal transduction; IEA.  
DR InterPro: IPR001610; PAC.  
DR InterPro: IPR000014; PAS.  
DR Pfam: PF00989; PAS; 1.  
DR SMART: SM00086; PAC; 2.  
DR SMART: SM00091; PAS; 2.  
DR PROSITE: PS50112; PAS; 2.  
SQ SEQUENCE 648 AA; 69995 MW; BBEFC744BC3F148B CRC64;  
Query March 35.0%; Score 573; DB 2; Length 648;  
Best Local Similarity 61.0%; Pred. No. 5.4e-39;  
Matches 125; Conservative 11; Mismatches 27; Indels 42; Gaps 5;  
QY 57 MRLTISYLRMRRLCAAGKGRATGRLPEGGGFRHGRHGRHGVKQC----- 109  
DB 1 MRLTISYLRMRRLCAAG-----EMNGVAGSGEPDLACVYKALEGF 40  
QY 110 ----QAPGPOVDLCSSSLIHNPTPGTNFSLIIGHISIPDFIHPCDQBELQALTPRPN 164  
DB 41 VMVLTAGDMVY-----LSENVSKHGLSGLLEIGHISIPDFIHPCDQBELQALTPROQT 94  
QY 165 LSKKKLEAPTEHRHPSLRMKSSTLTSGRTLNKAAATWVYLHSGHRAVYKPAQSPSPAGSP 224  
DB 95 LSRKRVLEPTEHRCSLRMKSSTLTSGRTLNKAAATWVYLHSGHRAVYKPAQSPSPAGSP 154  
QY 225 RSEPPLOCLVLCERAIPOLEPHDGA 249  
DB 155 DSEPPLOCLVLCERAI-----HPGS 175  
RESULT 10  
ID O9HAM5 PRELIMINARY; PRT; 199 AA.  
AC O9HAM5;  
DT 01-MAR-2001 (TReMBLrel. 16, Created)  
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)  
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)  
DE Hypothetical protein Faj11359.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=whole embryo;  
RX PubMed=14702039; DOI=10.1038/ng1285;  
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,  
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,  
RA Sekine M., Ohtsushi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,  
RA Yamamoto J., Saito K., Kawai T., Isono Y., Nakamura Y., Nagahara K.,  
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,  
RA Sudo H., Hosoiri T., Kaku T., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,  
RA Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M.,  
RA Niinomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,  
RA Tanai H., Kimata M., Watanabe M., Hiraoa S., Chiba Y., Ishida S.,  
RA Ono Y., Takiguchi S., Watanabe S., Yosida M., Houchi T., Kusano J.,  
RA Togashi K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,  
RA Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,  
RA Musasino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S.,  
RA Yoshihawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,  
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,  
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,  
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,  
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,

RA Fujimori Y., Komiyama M., Taehiro H., Tanigami A., Fujiwara T.,  
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y.,  
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,  
RA Ohtani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Sena T.,  
RA Matsunura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togaishi T., Oyama M., Hara H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isozaki T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RA Nat. Genet. 36:40-45(2004).  
RA EMBL: AK021421; BAB13819.1; -.  
DR HSPB: Q09814; IP97.  
SQ SEQUENCE 199 AA; 21598 MW; 646CB066A09B47C3 CRC64;  
Query March 31.7%; Score 518.5; DB 2; Length 199;  
Best Local Similarity 68.0%; Pred. No. 4.5e-35;  
Matches 104; Conservative 9; Mismatches 13; Indels 27; Gaps 2;  
QY 120 CSSSLIHNPTPGTNF-----SLEIGHISIPDFIHPCDQBELQ 156  
DB 14 CTSWLTREPSAPASAPTWTPRLSCASPSATCACTASAPQLEIGHISIPDFIHPCDQBELQ 73  
QY 157 DALTPRNLKKTLEAPTEHRHPSLRMKSSTLTSGRTLNKAAATWVYLHSGHRAVYKPPA 216  
DB 74 DALTPQTLSSRRVLEPTEHRCSLRMKSSTLTSGRTLNKAAATWVYLHSGHRAVYKPPA 133  
QY 217 QTSPPAGSPSEPPLOCLVLCERAIPOLEPHDGA 249  
DB 134 QTSPPAGSPSEPPLOCLVLCERAI-----HPGS 162  
RESULT 11  
ID HIFA CHICK STANDARD; PRT; 811 AA.  
AC 09YIB9;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (HIF1 alpha).  
GN Name=HIFA;  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RA Takahashi T.;  
RT "Molecular cloning and expression of an avian cDNA for hypoxia-  
inducible factor-1 alpha in embryonic ventricular myocytes.";  
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Functions as a master transcriptional regulator of the  
adaptive response to hypoxia. Binds to core DNA sequence 5'-  
[AG]CGTG-3' within the hypoxia response element (HRE) of target  
gene promoters. Activation requires recruitment of transcriptional  
coactivators (By similarity).  
CC -1- SUBUNIT: Efficient DNA binding requires heterodimerization of an  
alpha and a beta/ARNT subunit (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic in normoxia, nuclear  
translocation in response to hypoxia (By similarity).  
CC -1- DOMAIN: Contains two independent C-terminal transactivation  
domains, NTRAD and CTAD, which function synergistically. Their  
transcriptional activity is repressed by an intervening inhibitory  
domain (ID) (By similarity).  
CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
CC -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.  
CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB013746; BAA34234.2; --  
 DR HSSP; Q16665; 1L8C.  
 DR InterPro; IPR001092; HLH basic.  
 DR InterPro; IPR001321; Hypoxindf1A.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PRO1080; HYPOXIA1FLA.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PS0088; HLH; 1.  
 DR PROSITE; PS0112; PAS; 2.  
 KW Activator; DNA-binding; Nuclear protein; Repeat;  
 KW Transcription regulation.  
 FT DNA\_BIND 17 30 Basic motif.  
 FT DOMAIN 31 71 Helix-loop-helix motif.  
 FT DOMAIN 80 157 PAS 1.  
 FT DOMAIN 228 298 PAS 2.  
 FT DOMAIN 302 345 PAC.  
 FT DOMAIN 401 587 ODD.  
 FT DOMAIN 529 573 NTAD.  
 FT DOMAIN 576 785 ID.  
 FT DOMAIN 703 706 Nuclear localization signal (Potential).  
 FT DOMAIN 718 721 Nuclear localization signal (Potential).  
 FT DOMAIN 771 811 CTAD.  
 SQ DOMAIN 583 588 Poly-Ser.  
 SQ SEQUENCE 811 AA; 90542 MW; D14CDYFC98F064CB CRC64;  
 Query Match 30.7%; Score 503; DB 1; Length 811;  
 Best Local Similarity 49.8%; Pred. No. 4.6e-33;  
 Matches 119; Conservative 20; Mismatches 72; Indels 28; Gaps 5;  
 QY 9 RSNTELREKSRDARRSRQSEYLYOLANTLPARQVSAHLDAKASIMRLTISYLRMR 68  
 DB 12 RISSERRERKSRDARRSRQSEYLYOLANTLPARQVSAHLDAKASIMRLTISYLRMR 71  
 QY 69 LCAAGGKRGRRAT-----GRLLEPGGGRHGRHGRGRLGVCKCQQAQPGQSVDLCS 123  
 DB 72 LLDGSELTEANMEKELNCFYKALDGFVWVLSBDG-----DMYWS 113  
 QY 124 LIHNPTDG-TNPSLELIGHSIFDFIHPDQBELDALTTPRPNLSKKKLEAFTERHPSLRM 182  
 DB 114 ENWKKKGLTFP--DLTGHSVFDFTFPCDHELEKRLTHRGVPYKKGKQNTKTSFPLRM 171  
 QY 183 KSLTTSRGRTNLKAATWKLVLHCSGHMAYRPAQTSFAGSPRSEBPQLICVLCALP 241  
 DB 172 KCTLTSRGRTNLKAATWKLVLHCSGHMAYRPAQTSFAGSPRSEBPQLICVLCALP 228  
 RESULT 12  
 HIFA BOVIN STANDARD; PRT; 823 AA.  
 AC O9XTAS;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (HIF1 alpha).  
 GN Name=HIF1A;  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bos; Bovine; Bos  
 NC NCBI\_TaxID=9913;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Artery;  
 RX MEDLINE=99255430; PubMed=10320777; DOI=10.1016/S0167-4781(99)00048-2;  
 RA Hara S., Kobayashi C., Imura N.;  
 RT "Molecular cloning of cDNAs encoding hypoxia-inducible factor (HIF)-  
 RL 1alpha and -2alpha of bovine arterial endothelial cells.";  
 RL Biochim. Biophys. Acta 1445:237-243(1999).  
 CC -1- FUNCTION: Functions as a master transcriptional regulator of the  
 CC adaptive response to hypoxia. Under hypoxic conditions activates  
 CC the transcription of over 40 genes, including, erythropoietin,  
 CC glucose transporters, glycolytic enzymes, vascular endothelial  
 CC growth factor, and other genes whose protein products increase  
 CC oxygen delivery or facilitate metabolic adaptation to hypoxia.  
 CC Plays an essential role in embryonic vascularization, tumor  
 CC angiogenesis and pathophysiology of ischemic disease. Binds to  
 CC core DNA sequence 5'-(AG)CGTG-3' within the hypoxia response  
 CC element (HRE) of target gene promoters. Activation requires  
 CC recruitment of transcriptional coactivators such as CREBBP and  
 CC EP300. Activity is enhanced by interaction with both, NCOA1 or  
 CC NCOA2. Interaction with redox regulatory protein APEX seems to  
 CC activate CTAD and potentiates activation by NCOA1 and CREBBP (By  
 CC similarity).  
 CC -1- SUBUNIT: Efficient DNA binding requires heterodimerization of an  
 CC alpha and a beta/ARNT subunit. Binds to the TAZ-type 1 domains of  
 CC CREBBP and EP300. Interacts with NCOA1, NCOA2, APEX and HSP90 (By  
 CC similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic in normoxia, nuclear  
 CC translocation in response to hypoxia (By similarity).  
 CC -1- DOMAIN: Contains two independent C-terminal transactivation  
 CC domains, NTAD and CTAD, which function synergistically. Their  
 CC transcriptional activity is repressed by an intervening inhibitory  
 CC domain (ID) (By similarity).  
 CC -1- PTM: In normoxia, is hydroxylated on Pro-402 and Pro-564 in the  
 CC EGLN2/PHD2. EGLN3/PHD3 has also been shown to hydroxylate Pro-564.  
 CC The hydroxylated prolines promote interaction with VHL, initiating  
 CC rapid ubiquitination and subsequent proteasomal degradation. Under  
 CC hypoxia, proline hydroxylation is impaired and ubiquitination is  
 CC attenuated, resulting in stabilization (By similarity).  
 CC -1- PTM: In normoxia, is hydroxylated on Asn-800 by HIF1AN, thus  
 CC abrogating interaction with CREBBP and EP300 and preventing  
 CC transcriptional activation (By similarity).  
 CC -1- PTM: S-nitrosylated. All 15 free thiol groups are subjected to S-  
 CC nitrosylation in vitro, however not all thiol groups seem to be  
 CC nitrated in vivo (By similarity).  
 CC -1- PTM: Acetylation of Lys-532 by ARD1 increases interaction with VHL  
 CC and stimulates subsequent proteasomal degradation (By similarity).  
 CC -1- PTM: Requires phosphorylation for DNA-binding (By similarity).  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 CC -1- SIMILARITY: Contains 2 PAS (PER-ARNT-SIM) dimerization domains.  
 CC -1- SIMILARITY: Contains 1 PAS-associated C-terminal (PAC) domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AB018398; BAA78675.1; --  
 DR HSSP; Q16665; 1L8C.  
 DR InterPro; IPR001092; HLH basic.  
 DR InterPro; IPR001321; Hypoxindf1A.  
 DR InterPro; IPR001610; PAC.  
 DR InterPro; IPR000014; PAS.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PRO1080; HYPOXIA1FLA.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.

DR PROSITE; PSS0888; HLH; 1.  
 DR PROSITE; PSS0112; PAS; 2.  
 KW Acetylation; Activator; DNA-binding; Hydroxylation; Nuclear protein;  
 KW Phosphorylation; Repeat; S-nitrosylation; Transcription regulation.  
 FT DOMAIN 718 721  
 FT DNA\_BIND 17 30  
 FT DOMAIN 31 71  
 FT DOMAIN 85 158  
 FT DOMAIN 228 298  
 FT DOMAIN 302 345  
 FT DOMAIN 401 600  
 FT DOMAIN 531 575  
 FT DOMAIN 576 782  
 FT DOMAIN 783 823  
 FT DOMAIN 715 718  
 FT MOD\_RES 90 90  
 FT MOD\_RES 139 139  
 FT MOD\_RES 173 173  
 FT MOD\_RES 194 194  
 FT MOD\_RES 210 210  
 FT MOD\_RES 219 219  
 FT MOD\_RES 224 224  
 FT MOD\_RES 255 255  
 FT MOD\_RES 334 334  
 FT MOD\_RES 337 337  
 FT MOD\_RES 359 359  
 FT MOD\_RES 402 402  
 FT MOD\_RES 520 520  
 FT MOD\_RES 532 532  
 FT MOD\_RES 564 564  
 FT MOD\_RES 755 755  
 FT MOD\_RES 777 777  
 FT MOD\_RES 797 797  
 FT MOD\_RES 800 800  
 SQ SEQUENCE 823 AA; 92127 MW; 12674E67A61B1A1 CRC64;

Query Match 30.5%; Score 499; DB 1; Length 823;  
 Best Local Similarity 46.6%; Pred. No. 1e-32;  
 Matches 115; Conservative 26; Mismatches 58; Indels 48; Gaps 6;

QY 11 NTELRKESKRDAAARRSOETEVLYOLAHTLPARGVSAHLDKASIMRLTISYLRMRLC 70  
 Db 14 SSKRRKESKRDAAARRSKSESEVFYELAHQLPLPHNVSSHLDKASVWRLTISYLRVKLL 73  
 QY 71 AAGKGRATGRLLPEPGGFRGTHRGHGLPVKCCQA-----PGQSVYDLC 120  
 Db 74 DAG-----DLTIDEMKAQNNCFYLKALDGFVWLTD 105  
 QY 121 SSSLIH-----NPTPG-TNPSLELIGHSIPFTHPCDSELDLTPRNLSSKKLEAPF 174  
 Db 106 DGDMIYISDNVKNKMGITQF--ELTGHSVDFTHPCDHEMRMLTHRNGLVKKGEQNT 163  
 QY 175 ERHPSLMMKSTLTSGRTINLKAATWKLHCSGMRAYKPAQTSFAGSPRSEPPLOCTV 234  
 Db 164 QRSFPLMKCTLTSGRTINLKAATWKLHCSGMRAYKPAQTSFAGSPRSEPPLOCTV 221  
 QY 235 LICEAIP 241  
 Db 222 LICEPIP 228  
 RESULT 13  
 Q6IV47  
 ID Q6IV47 PRELIMINARY; PRT; 823 AA.  
 AC Q6IV47;  
 DT 05-JUN-2004 (TRENBLrel. 27, Created)  
 DT 05-JUN-2004 (TRENBLrel. 27, Last sequence update)  
 DE Hypoxia-inducible factor-1a.  
 GN Name=HIF-1A;  
 OS Bos mutus graminis (Yak).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Mammalia;

OC Bovinae; Bos.  
 OX NCBI\_TaxID=30521;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Dolt K.S., Qadar Pasha M.A.;  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
 DR EMBL; AY621118; AY621118.1; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0004871; F:signal transducer activity; IEA.  
 DR GO; GO:0003700; F:transcription factor activity; IEA.  
 DR GO; GO:0006355; P:regulation of transcription; DNA-dependent; IEA.  
 DR GO; GO:0007165; P:signal transduction; IEA.  
 DR InterPro; IPR001092; HLH\_Basic.  
 DR InterPro; IPR001321; Hypoxindf1A.  
 DR InterPro; IPR001610; PAC.  
 DR Pfam; PF00010; HLH; 1.  
 DR Pfam; PF00785; PAC; 1.  
 DR Pfam; PF00989; PAS; 2.  
 DR PRINTS; PR01080; HYPOXIA1F1A.  
 DR SMART; SM00353; HLH; 1.  
 DR SMART; SM00086; PAC; 1.  
 DR SMART; SM00091; PAS; 2.  
 DR PROSITE; PSS0888; HLH; 1.  
 DR PROSITE; PSS0112; PAS; 2.  
 SQ SEQUENCE 823 AA; 92128 MW; A6E388E4FEA15705 CRC64;

Query Match 30.5%; Score 499; DB 2; Length 823;  
 Best Local Similarity 46.6%; Pred. No. 1e-32;  
 Matches 115; Conservative 26; Mismatches 58; Indels 48; Gaps 6;

QY 11 NTELRKESKRDAAARRSOETEVLYOLAHTLPARGVSAHLDKASIMRLTISYLRMRLC 70  
 Db 14 SSKRRKESKRDAAARRSKSESEVFYELAHQLPLPHNVSSHLDKASVWRLTISYLRVKLL 73  
 QY 71 AAGKGRATGRLLPEPGGFRGTHRGHGLPVKCCQA-----PGQSVYDLC 120  
 Db 74 DAG-----DLTIDEMKAQNNCFYLKALDGFVWLTD 105  
 QY 121 SSSLIH-----NPTPG-TNPSLELIGHSIPFTHPCDSELDLTPRNLSSKKLEAPF 174  
 Db 106 DGDMIYISDNVKNKMGITQF--ELTGHSVDFTHPCDHEMRMLTHRNGLVKKGEQNT 163  
 QY 175 ERHPSLMMKSTLTSGRTINLKAATWKLHCSGMRAYKPAQTSFAGSPRSEPPLOCTV 234  
 Db 164 QRSFPLMKCTLTSGRTINLKAATWKLHCSGMRAYKPAQTSFAGSPRSEPPLOCTV 221  
 QY 235 LICEAIP 241  
 Db 222 LICEPIP 228  
 RESULT 14  
 HIFA\_RAT  
 ID HIFA\_RAT STANDARD; PRT; 825 AA.  
 AC O35800; Q9WTU9;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DE Hypoxia-inducible factor 1 alpha (HIF-1 alpha) (HIF1 alpha).  
 GN Name=Hif1a;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Wistar; TISSUE=Hepatocytes;  
 RC MEDLINE=1134367; PubMed=11237857; DOI=10.1042/0264-6021:3540531;  
 RA Kietzmann T., Cornesse Y., Brechtel K., Modaresi S., Jungermann K.;  
 RT "Periveneous expression of the mRNA of the three hypoxia-inducible factor a-subunits HIF-1a, HIF2a and HIF3a in rat liver.";

Biochem. J. 354:531-537(2001).

RN  
RN  
RN  
SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Kidney;  
RX MEDLINE=21417706; PubMed=11526200;  
RA Zou A.-P., Yang Z.-Z., Li P.-L., Cowley A.W. Jr.;  
RT "Oxygen-dependent expression of hypoxia-inducible factor-1alpha in renal medullary cells of rats.";  
RL Physiol. Genomics 6:159-168(2001).  
CC -1- FUNCTION: Functions as a master transcriptional regulator of the adaptive responses to hypoxia. Under hypoxic conditions activates the transcription of over 40 genes, including, erythropoietin, glucose transporters, glycolytic enzymes, vascular endothelial growth factor, and other genes whose protein products increase oxygen delivery or facilitate metabolic adaptation to hypoxia. Plays an essential role in embryonic vasculatization, tumor angiogenesis and pathophysiology of ischemic disease. Binds to core DNA sequence 5'-[AG]CGTG-3' within the hypoxia response element (HRE) of target gene promoters. Activation requires recruitment of transcriptional coactivators such as CREBBP and EP300. Activity is enhanced by interaction with both, NCOA1 or NCOA2. Interaction with relox regulatory protein APX seems to activate CTAD and potentiates activation by NCOA1 and CREBBP (By similarity).  
CC -1- SUBUNIT: Efficient DNA binding requires heterodimerization of an alpha and a beta/ARNT subunit. Binds to the T12-type 1 domains of CREBBP and EP300. Interacts with NCOA1, NCOA2, APX and HSP90 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic in normoxia, nuclear translocation in response to hypoxia (By similarity).  
CC -1- TISSUE SPECIFICITY: Expressed in the kidney, higher expression is seen in the renal medulla than in the cortex. Expressed also in the perivascular zone of the liver.  
CC -1- DOMAIN: Contains two independent C-terminal transactivation domains, NTAD and CTAD, which function synergistically. Their transcriptional activity is repressed by an intervening inhibitory domain (ID) (By similarity).  
CC -1- PTM: In normoxia, is hydroxylated on Pro-402 and Pro-563 in the oxygen-dependent degradation domain (ODD) by EGLN1/PHD1 and EGLN2/PHD2. EGLN3/PHD3 has also been shown to hydroxylate Pro-563. The hydroxylated prolines promote interaction with VHL, initiating rapid ubiquitination and subsequent proteasomal degradation. Under hypoxia, proline hydroxylation is impaired and ubiquitination is attenuated, resulting in stabilization (By similarity).  
CC -1- PTM: In normoxia, is hydroxylated on Asn-802 by HIF1AN, thus abrogating interaction with CREBBP and EP300 and preventing transcriptional activation (By similarity).  
CC -1- PTM: S-nitrosylated. All free thiol groups are subjected to S-nitrosylation in vitro, however not all thiol groups seem to be nitrosylated in vivo (By similarity).  
CC -1- PTM: Acetylation of Lys-531 by ARBD increases interaction with VHL and stimulates subsequent proteasomal degradation (By similarity).  
CC -1- PTM: Phosphorylation is required for DNA binding (By similarity).  
CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.  
CC -1- SIMILARITY: Contains 2 PAS (PBR-ARNT-SIM) dimerization domains.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announce/](http://www.isb-sib.ch/announce/or send an email to licenses@isb-sib.ch) or send an email to licenses@isb-sib.ch).  
CC  
CC  
CC EMBL; Y09507; CAA70701.1; -  
DR EMBL; AF057308; AAD24413.1; -  
DR HSSP; Q16655; ILBC.  
DR TRASNFPAC; T05461; -  
DR INTERPRO; IPR001921; HLH\_basic.  
DR INTERPRO; IPR001321; HypoxIndf1A.  
DR INTERPRO; IPR001610; PAC.  
DR INTERPRO; IPR000014; PAS.

Query Match	Similarity	Score	DB 1	Length	825
Best Local Similarity	46.7%	Pred. No. 1.56-33;			
Matches 113;	Conservative 29;	Mismatches 72;	Indels 28;	Gaps 5;	
QY	6 QVRNSNTELRKESKRDARSRRSOETEVLYQLAHTLPPFANGVSAHLDKASIMRLTISYLR	65			
DB	9 EKKMSSESRKESKRDARSRRSSESVFELAHQPLPNVNSHILDKASVMRLTISYLR	68			
QY	66 MHRICAG-----CKRGATGRLLRBERGGFRGTHRGHNGLPVCKCQDAPRGQSYDLC	120			
DB	69 VRKLLDGDLDIEDMKQNNCFYLKAPDFVWVLTDDG-----DMT	110			
QY	121 SSSLINHPTEPG-TWFSLELIGHSIFDPIHCDOEELDALTPRPNLSSKKKLEAPTERHFS	179			
DB	111 YISNNVKKMGLOTF--ELTGHVSFDPTTHCDHEHMEMLTHNGPRARKKEQNTQTSFF	168			
QY	180 LRKMSYLLTSRGRTLNLLKAATWKLVLKSGHMRAYKPPAQTS PAPS PRSEPPQLVILI	239			
DB	169 LRKMSYLLTSRGRTLNLLKAATWKLVLKSGHMRAYKPPAQTS PAPS PRSEPPQLVILI	226			
QY	240 IP 241				
DB	227 IP 228				



```
RESULT 15
Q6EM13 PRELIMINARY; PRT; 489 AA.
ID Q6EM13
AC Q6EM13;
DT 25-OCT-2004 (TRENBLREL. 28, Created)
DT 25-OCT-2004 (TRENBLREL. 28, Last sequence update)
DT 25-OCT-2004 (TRENBLREL. 28, Last annotation update)
DE Hypoxia-inducible factor 1 alpha (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX PubMed=15162502; DOI=10.1002/dvdy.20049;
RA Sipe C.W., Gruber E.J., Saha M.S.;
RT "Short upstream region drives dynamic expression of hypoxia-inducible
factor 1alpha during Xenopus development.";
RL Dev. Dyn. 230:229-238(2004).
CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
DR EMBL; AY189821; AAC72733.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR01092; HLH basic.
DR InterPro; IPR01321; Hypoxindf1A.
DR InterPro; IPR01067; Nuc_translocat.
DR InterPro; IPR01610; PAC.
DR InterPro; IPR00014; PAS.
DR Pfam; PF00010; HLH; 1.
DR Pfam; PF00785; PAC; 1.
DR Pfam; PF00989; PAS; 2.
DR PRINTS; PR01080; HYPOXIA1FLA.
DR PRINTS; PR00785; NCTRNLOCATR.
DR SMART; SM00353; HLH; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 2.
DR PROSITE; PS50888; HLH; 1.
DR PROSITE; PS50112; PAS; 2.
FT NON_TER 1
FT NON_TER 489
FT NON_TER 1
SQ SEQUENCE 489 AA; 55613 MW; 39E1ACA3E6AF766 CRC64;

Query Match 30.3%; Score 495; DB 2; Length 489;
Best Local Similarity 48.6%; Pred. No. 1.2e-32;
Matches 114; Conservative 27; Mismatches 71; Indels 18; Gaps 6;

QY 13 ELRREKSRDAARSRROETEVLYQLAHTLPPARGVSAHLDKASIMRLTISYLRMRLCAA 72
DB 1 ERREKSRDARCRSRKSESVFYELSHLPLPHNVSSHLDKASIMRLAISYLRRLDLA 60
QY 73 GGRKGRATGRLLPBGPGFRHGRHGLPVGKCOAPGPQSVYDLCSSSLINPTPG- 131
DB 61 GDLDEETD---LDQLNCF---YLKALEGFVLVLTREG-----DWIYSENVMKCMGL 107
QY 132 TNFSLTELIGHSIFPFIRPCDEELQDALTPRPNLSSKKKLKAPTERHPSLRKSTLTSRGR 191
DB 108 TGF--ELTGHSVFPFTRPCDEELREMLTFRNGPAKKGKEQITERSFFLRKCTVTSRGR 165
QY 192 TLNLKAATWKLVCSGHMRAYKPPAQTSPPAGSPRSEPPLOCLVLICEAIP 241
DB 166 TVNIKSAITWKLVCYGHMRVYDNNANQKHG--YKRPVKCMVVICEPI 213
```

Search completed: June 15, 2005, 16:00:09  
Job time : 176 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 16:00:19 ; Search time 157 Seconds  
(without alignments)  
749.578 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALGGLQVRNTELRKEKSR.....TESSLPSWVLMALNRKNCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1710399 seqs, 38333425 residues

Total number of hits satisfying chosen parameters: 1710399

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA: \*  
1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubppaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubppaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubppaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubppaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/1/pubppaa/US10E\_PUBCOMB.pep:\*  
18: /cgn2\_6/ptodata/1/pubppaa/US10F\_PUBCOMB.pep:\*  
19: /cgn2\_6/ptodata/1/pubppaa/US11A\_PUBCOMB.pep:\*  
20: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep:\*  
21: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*  
22: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	307	9	US-09-896-791B-3
2	1278	78.1	324	15	US-10-307-828A-20
3	795	48.6	705	14	US-10-154-386-2
4	498	30.4	827	10	US-09-919-039-149
5	498	30.4	827	14	US-10-247-671-137
6	495	30.3	623	10	US-09-967-388-2
7	495	30.3	735	16	US-10-032-361-2
8	495	30.3	826	9	US-09-922-958-4
9	495	30.3	826	9	US-09-833-790-235
10	495	30.3	826	9	US-09-736-457-330
11	495	30.3	826	9	US-09-902-941-330

12	495	30.3	826	9	US-09-849-626-330	Sequence 330, App
13	495	30.3	826	10	US-09-967-388-4	Sequence 4, Appl1
14	495	30.3	826	10	US-09-476-300-330	Sequence 330, App
15	495	30.3	826	13	US-10-028-158-23	Sequence 20, Appl
16	495	30.3	826	13	US-10-101-812-10	Sequence 10, Appl
17	495	30.3	826	14	US-10-101-662A-9	Sequence 9, Appl1
18	495	30.3	826	14	US-10-101-816-2	Sequence 2, Appl1
19	495	30.3	826	14	US-10-101-816-5	Sequence 5, Appl1
20	495	30.3	826	14	US-10-101-816-6	Sequence 6, Appl1
21	495	30.3	826	14	US-10-101-816-7	Sequence 7, Appl1
22	495	30.3	826	14	US-10-017-754-330	Sequence 330, App
23	495	30.3	826	14	US-10-115-987B-14	Sequence 14, Appl
24	495	30.3	826	14	US-10-287-670-9	Sequence 9, Appl1
25	495	30.3	826	14	US-10-113-872-330	Sequence 330, App
26	495	30.3	826	14	US-10-423-419-2	Sequence 2, Appl1
27	495	30.3	826	15	US-10-283-017-330	Sequence 330, App
28	495	30.3	826	16	US-10-032-361-1	Sequence 1, Appl1
29	495	30.3	826	16	US-10-854-483-2	Sequence 2, Appl1
30	495	30.3	826	16	US-10-854-483-5	Sequence 5, Appl1
31	495	30.3	826	16	US-10-854-483-6	Sequence 6, Appl1
32	495	30.3	826	16	US-10-854-483-7	Sequence 7, Appl1
33	495	30.3	826	16	US-10-859-935-10	Sequence 10, Appl
34	495	30.3	826	16	US-10-859-935-23	Sequence 23, Appl
35	495	30.3	826	17	US-10-901-583-18	Sequence 18, Appl
36	491	30.0	823	14	US-10-205-342-13	Sequence 13, Appl
37	467	28.5	870	14	US-10-101-816-4	Sequence 4, Appl1
38	467	28.5	870	16	US-10-854-483-4	Sequence 4, Appl1
39	466	28.5	485	14	US-10-121-235-6	Sequence 6, Appl1
40	466	28.5	870	14	US-10-121-235-6	Sequence 6, Appl1
41	281	17.2	1507	14	US-10-080-608A-37	Sequence 37, Appl
42	281	17.2	1507	15	US-10-370-685-126	Sequence 126, App
43	278.5	17.0	901	16	US-10-408-765A-1259	Sequence 1259, App
44	276.5	16.9	903	16	US-10-408-765A-1332	Sequence 1332, App
45	256	15.6	570	9	US-09-923-684-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1  
US-09-896-791B-3  
Sequence 3, Application US/09896791B  
Patent No. US20020165140A1  
GENERAL INFORMATION  
APPLICANT: Bertkenstam, Anders  
APPLICANT: Bertilsson, Gran  
APPLICANT: Poellinger, Lorenz  
TITLE OF INVENTION: SCREENING METHODS  
FILE REFERENCE: 13425-040001  
CURRENT APPLICATION NUMBER: US/09/896, 791B  
CURRENT FILING DATE: 2001-06-29  
PRIOR APPLICATION NUMBER: 60/217, 570  
PRIOR FILING DATE: 2000-07-12  
PRIOR APPLICATION NUMBER: SE 0002551-0  
PRIOR FILING DATE: 2000-07-06  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 307  
TYPE: PRT  
ORGANISM: Mus musculus  
US-09-896-791B-3  
Query Match 100.0%; Score 1636; DB 9; Length 307;  
Best Local Similarity 100.0%; Pred. No. 6.3e-146; Indels 0; Gaps 0;  
Matches 307; Conservative 0; Mismatches 0;  
QY 1 MALGGLQVRNTELRKEKSRDARSRSQSTEVLYQLAHTLPFARGVSAHLDKASINRLT 60  
DB 1 MALGGLQVRNTELRKEKSRDARSRSQSTEVLYQLAHTLPFARGVSAHLDKASINRLT 60  
QY 61 ISYIRMRLCAGGKGRATGRLLPBGPGFRRGTHRRGSHGLPVGCGQAPGQSVYDLC 120

```

Db      61 ISYLRMRRLCAAGKGRGATGRLLPEGPGGFRHGTNRGRHGLPVGKCOQAPGQSVDC 120
Qy      121 SSSLINHPPTGNTFSLIGHISIFDFIHPDQBELQDALTPRPNLSSKKLEAPTERHFSL 180
Db      121 SSSLINHPPTGNTFSLIGHISIFDFIHPDQBELQDALTPRPNLSSKKLEAPTERHFSL 180
Qy      181 RMKSTLTSRGTINLKAATWVHLHCSGHRAYKPPAOTSPPGSRSEPPLOCLVLI CEAI 240
Db      181 RMKSTLTSRGTINLKAATWVHLHCSGHRAYKPPAOTSPPGSRSEPPLOCLVLI CEAI 240
Qy      241 POLPFHGGATLGLPOEKTPISTLFTPLMKALLCLVKMPVQVLOGKGTESLPSVWLAL 300
Db      241 POLPFHGGATLGLPOEKTPISTLFTPLMKALLCLVKMPVQVLOGKGTESLPSVWLAL 300
Qy      301 NRKNCPG 307
Db      301 NRKNCPG 307

RESULT 2
US-10-307-928A-20
; Sequence 20, Application US/10307928A
; Publication No. US20030229016A1
; GENERAL INFORMATION:
; APPLICANT: Alsebrook, John P.
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Catterton, Elna
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gorman, Linda
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Ji, Weizhen
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Li, Li
; APPLICANT: Paturajan, Meera
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Vermet, Corine A.M.
; APPLICANT: Voess, Edward Z.
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS C
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 24102-502D
; CURRENT APPLICATION NUMBER: US/10/307,928A
; PRIOR FILING DATE: 2002-12-02
; PRIOR APPLICATION NUMBER: 60/341,477
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/341,540
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/342,592
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 60/344,903
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/373,288
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/380,981
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: 60/381,495
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/383,744
; PRIOR FILING DATE: 2002-05-28
; PRIOR APPLICATION NUMBER: 60/384,024
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: 60/401,788
; PRIOR FILING DATE: 2002-08-07
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: Curseseglist version 0.1
; SEQ ID NO 20
; LENGTH: 324
; TYPE: PRT

```

```

; ORGANISM: Homo sapiens
US-10-307-928A-20
Query Match 78.1%; Score 1278; DB 15; Length 324;
Best Local Similarity 79.9%; Pred. No. 4,6e-112;
Matches 250; Conservative 14; Mismatches 41; Indels 8; Gaps 3;

Qy      1 MALGLQVRNNTLRKESRDARSRSQTEVLYOLAHTLPFARGVSAHLDAKASIMRLT 60
Db      1 MALGLQVRNNTLRKESRDARSRSQTEVLYOLAHTLPFARGVSAHLDAKASIMRLT 60
Qy      61 ISYLRMRRLCAAGKGRGATGRLLPEGPGGFRHGTNRGRHGLPVGKCOQAPGQSVDC 120
Db      61 ISYLRMRRLCAAGKGRGATGRLLPEGPGGFRHGTNRGRHGLPVGKCOQAPGQSVDC 120
Qy      121 SSSLINHPPTGNTFSLIGHISIFDFIHPDQBELQDALTPRPNLSSKKLEAPTERHFSL 180
Db      121 SSSLINHPPTGNTFSLIGHISIFDFIHPDQBELQDALTPRPNLSSKKLEAPTERHFSL 180
Qy      181 RMKSTLTSRGTINLKAATWVHLHCSGHRAYKPPAOTSPPGSRSEPPLOCLVLI CEAI 240
Db      181 RMKSTLTSRGTINLKAATWVHLHCSGHRAYKPPAOTSPPGSRSEPPLOCLVLI CEAI 240
Qy      241 POLPFHGGATLGLPOEKTPISTLFTPLMKALLCLVKMPVQVLOGKGTESLPS 294
Db      241 PSFLSTGGATLGLPOEKA--SPILSPYVCGHCSLCFVEMPDIQAMGSDPSLLPL 298
Qy      295 WVLALNRKNCPG 307
Db      295 EVWMALNGNCAG 311

RESULT 3
US-10-154-386-2
; Sequence 2, Application US/10154386
; Publication No. US20030026793A1
; GENERAL INFORMATION:
; APPLICANT: Angiogene Inc.
; APPLICANT: Guy, Louis-Georges
; TITLE OF INVENTION: HIPPOXIA INDUCING FACTORS AND USES THEREOF FOR INDUCING ANGIOGENESIS
; TITLE OF INVENTION: IMPROVING MUSCULAR FUNCTIONS
; FILE REFERENCE: 5600-81
; CURRENT APPLICATION NUMBER: US/10/154,386
; PRIOR FILING DATE: 2002-05-23
; PRIOR APPLICATION NUMBER: US 60/292,630
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/354529
; PRIOR FILING DATE: 2002-02-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 705
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-154-386-2
Query Match 48.6%; Score 795; DB 14; Length 705;
Best Local Similarity 67.6%; Pred. No. 5,6e-66;
Matches 171; Conservative 11; Mismatches 29; Indels 42; Gaps 5;

Qy      9 RSTTELRKESRDARSRSQTEVLYOLAHTLPFARGVSAHLDAKASIMRLTISYLRMR 68
Db      3 RSTTELRKESRDARSRSQTEVLYOLAHTLPFARGVSAHLDAKASIMRLTISYLRMR 62
Qy      69 LCAAGKGRGATGRLLPEGPGGFRHGTNRGRHGLPVGKCOQAPGQSV 117
Db      63 LCAAG-----EMNVGAGGEPLDACYLKALGFVWVLTABGDMAY 102
Qy      118 DLSSSLINHPPTGNTFSLIGHISIFDFIHPDQBELQDALTPRPNLSSKKLEAPTER 176
Db      103 -----LSENVSGHGLGSELEIGHISIFDFIHPDQBELQDALTPQOTLSRRVBAFTER 156
Qy      177 HFSLRMKSTLTSRGTINLKAATWVHLHCSGHRAYKPPAOTSPPGSRSEPPLOCLVLI 236

```

```
Db 157 CFSLRMKSTLTSRGRITNLKAAATWVNLCSGHRMAKYPAPQTSPPAGSPSEPPLOCPLI 216
Qy 237 CEATPOLPFHDGA 249
Db 217 CEALP----HPGS 225
```

## RESULT 4

```
US-09-919-039-149
; Sequence 149, Application US/09919039
; Publication No. US20030108871A1
; GENERAL INFORMATION:
; APPLICANT: Kaeber, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES
; FILE REFERENCE: PA-0035 US
; CURRENT APPLICATION NUMBER: US/09/919, 039
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: 60/222,113
; NUMBER OF SEQ ID NOS: 401
; SOFTWARE: PERL Program
; SEQ ID NO 149
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030108871A1 1250434CD1
US-09-919-039-149
```

```
Query Match 30.4%; Score 498; DB 10; Length 827;
Best Local Similarity 46.2%; Pred. No. 8e-38;
Matches 115; Conservative 26; Mismatches 60; Indels 48; Gaps 6;
```

```
Qy 9 RSNTELRKESRDAAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMR 68
Db 13 RISERREKESRDAAARRSKESVFEYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRK 72
Qy 69 LCAAGKRGATGRLLPBGPGFGRHGRHGLPVGKCOQA-----PGQSVLD 118
Db 73 LLDG-----DLIDEDMKAKMNCFTYKALDGFVWVL 104
Qy 119 LCSSSLIH-----NPTPG-TNFSLELIGHISFDPIHPCDQELQDALTPRNLSSKKLEA 172
Db 105 TDGDMITVYISDNVKNKMGLTQF--ELTGHSVFDPFHCDHEKEMLTNRNGLVKKGEQ 162
Qy 173 PTERHPSLRMKSTLTSRGRITNLKAAATWVNLCSGHRMAKYPAPQTSPPAGSPSEPPLOC 232
Db 163 NTQGSFPLRMKCTLTSRGRITNLSKATWVNLCHTGHIVYDTNSNPQCG--YKKPPMTC 220
Qy 233 LVLICEAIP 241
Db 221 LVLICEPIP 229
```

## RESULT 5

```
US-10-247-671-137
; Sequence 137, Application US/10247671
; Publication No. US20030194721A1
; GENERAL INFORMATION:
; APPLICANT: Mikita, Thomas
; APPLICANT: Shiffman, Dov
; APPLICANT: Porter, Gordon, J.
; APPLICANT: Kaeber, Matthew R.
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
; FILE REFERENCE: PA-0050 US
; CURRENT APPLICATION NUMBER: US/10/247, 671
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: 60/323,784
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: PERL Program
```

```
; SEQ ID NO 137
; LENGTH: 827
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030194721A1 1250434CD1
US-10-247-671-137
```

```
Query Match 30.4%; Score 498; DB 14; Length 827;
Best Local Similarity 46.2%; Pred. No. 8e-38;
Matches 115; Conservative 26; Mismatches 60; Indels 48; Gaps 6;
```

```
Qy 9 RSNTELRKESRDAAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMR 68
Db 13 RISERREKESRDAAARRSKESVFEYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRK 72
Qy 69 LCAAGKRGATGRLLPBGPGFGRHGRHGLPVGKCOQA-----PGQSVLD 118
Db 73 LLDG-----DLIDEDMKAKMNCFTYKALDGFVWVL 104
Qy 119 LCSSSLIH-----NPTPG-TNFSLELIGHISFDPIHPCDQELQDALTPRNLSSKKLEA 172
Db 105 TDGDMITVYISDNVKNKMGLTQF--ELTGHSVFDPFHCDHEKEMLTNRNGLVKKGEQ 162
Qy 173 PTERHPSLRMKSTLTSRGRITNLKAAATWVNLCSGHRMAKYPAPQTSPPAGSPSEPPLOC 232
Db 163 NTQGSFPLRMKCTLTSRGRITNLSKATWVNLCHTGHIVYDTNSNPQCG--YKKPPMTC 220
Qy 233 LVLICEAIP 241
Db 221 LVLICEPIP 229
```

## RESULT 6

```
US-09-967-388-2
; Sequence 2, Application US/09967388
; Publication No. US20030103956A1
; GENERAL INFORMATION:
; APPLICANT: JEFFREY M. ARBETT
; TITLE OF INVENTION: USE OF HIF-1ALPHA VARIANTS TO ACCELERATE
; FILE REFERENCE: UC077,001A
; CURRENT APPLICATION NUMBER: US/09/967,388
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 623
; TYPE: PRT
; ORGANISM: HUMAN
US-09-967-388-2
```

```
Query Match 30.3%; Score 495; DB 10; Length 623;
Best Local Similarity 46.2%; Pred. No. 1.e-37;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
Qy 11 NTELRKESRDAAARRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMR 70
Db 14 SSERREKESRDAAARRSKESVFEYELAHQLPLPHNVSSHLDKASVMRLTISYLRVRK 73
Qy 71 AAGKRGATGRLLPBGPGFGRHGRHGLPVGKCOQA-----PGQSVLD 120
Db 74 DAG-----DLIDEDMKAKMNCFTYKALDGFVWVLTD 105
Qy 121 SSSSLIH-----NPTPG-TNFSLELIGHISFDPIHPCDQELQDALTPRNLSSKKLEA 174
Db 106 DGMITVYISDNVKNKMGLTQF--ELTGHSVFDPFHCDHEKEMLTNRNGLVKKGEQ 163
Qy 175 ERHPSLRMKSTLTSRGRITNLKAAATWVNLCSGHRMAKYPAPQTSPPAGSPSEPPLOC 234
Db 164 QRSFPLRMKCTLTSRGRITNLSKATWVNLCHTGHIVYDTNSNPQCG--YKKPPMTC 221
```



```
US-09-736-457-330
; Sequence 330, Application US/09736457
; Patent No. US2002016837A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 330
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-736-457-330
```

Query March 30.3%; Score 495; DB 9; Length 826;

Best Local Similarity 46.2%; Pred. No. 1.5e-37;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

```
QY 11 NTELRKESRDAAARRSRSQETEVLYQLAHTLPFARGVSAHLDRKASIMRLTISYLRMRLC 70
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERKESKSDAARRSRSKSESEVFEYELAHQLPPLHNVS SHLDKASVRLTISYLRVRL 73
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 71 AAGKRGATGRLLPEPGGFRHGRHGLPVGKCOA-----PGQSYDLC 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLIEDMKAKQNCFFYLKALDGFVWLTD 105
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 SSSLIH-----NPTPG-TNFSLELIGHSIFDPIPCQOELOALTPRPNISSKKLEAPT 174
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKYMGLTF--ELTGHSVDFPTHPCHEEREMLTNRNGLVKKGKQNT 163
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 175 ERHFSLMKSTLTSRGRTNLKATWVKVLCSCGHRAYKPPAOTS PAPSRSBPLOCLV 234
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 QRSFPLMKCTLSRGRTNMKSAITWVKVLCGHIHYDTNNSNPQCG--YKKEPMTCLV 221
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 235 LICEAIP 241
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 222 LICEPIP 228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 11
US-09-902-941-330
; Sequence 330, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Wang, Tongtong
; APPLICANT: Watanabe, Yoshihiro
; APPLICANT: Johnson, Jeffrey C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Marnerakis, Margarita
; APPLICANT: Carter, Darick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: McNabb, Andria
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C17
; CURRENT APPLICATION NUMBER: US/09/902,941
; CURRENT FILING DATE: 2001-07-10
```

```
; NUMBER OF SEQ ID NOS: 2002
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 330
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-902-941-330
```

Query March 30.3%; Score 495; DB 9; Length 826;

Best Local Similarity 46.2%; Pred. No. 1.5e-37;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

```
QY 11 NTELRKESRDAAARRSRSQETEVLYQLAHTLPFARGVSAHLDRKASIMRLTISYLRMRLC 70
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERKESKSDAARRSRSKSESEVFEYELAHQLPPLHNVS SHLDKASVRLTISYLRVRL 73
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 71 AAGKRGATGRLLPEPGGFRHGRHGLPVGKCOA-----PGQSYDLC 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLIEDMKAKQNCFFYLKALDGFVWLTD 105
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 SSSLIH-----NPTPG-TNFSLELIGHSIFDPIPCQOELOALTPRPNISSKKLEAPT 174
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKYMGLTF--ELTGHSVDFPTHPCHEEREMLTNRNGLVKKGKQNT 163
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 175 ERHFSLMKSTLTSRGRTNLKATWVKVLCSCGHRAYKPPAOTS PAPSRSBPLOCLV 234
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 QRSFPLMKCTLSRGRTNMKSAITWVKVLCGHIHYDTNNSNPQCG--YKKEPMTCLV 221
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 235 LICEAIP 241
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 222 LICEPIP 228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
RESULT 12
US-09-849-626-330
; Sequence 330, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
; APPLICANT: Bangur, Chaitanya
; APPLICANT: Fanger, Gary
; APPLICANT: Wang, Aijun
; APPLICANT: Wang, Tongtong
; APPLICANT: McNeel, Anne
; APPLICANT: Switzer, Patricia
; APPLICANT: Clapper, Jonathan
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C16
; CURRENT APPLICATION NUMBER: US/09/849,626
; CURRENT FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 1926
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 330
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-849-626-330
```

Query March 30.3%; Score 495; DB 9; Length 826;

Best Local Similarity 46.2%; Pred. No. 1.5e-37;  
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;

```
QY 11 NTELRKESRDAAARRSRSQETEVLYQLAHTLPFARGVSAHLDRKASIMRLTISYLRMRLC 70
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERKESKSDAARRSRSKSESEVFEYELAHQLPPLHNVS SHLDKASVRLTISYLRVRL 73
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 71 AAGKRGATGRLLPEPGGFRHGRHGLPVGKCOA-----PGQSYDLC 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLIEDMKAKQNCFFYLKALDGFVWLTD 105
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 SSSLIH-----NPTPG-TNFSLELIGHSIFDPIPCQOELOALTPRPNISSKKLEAPT 174
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKYMGLTF--ELTGHSVDFPTHPCHEEREMLTNRNGLVKKGKQNT 163
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```



Oy	235	LICRAIP	241
Db	222	LICRPIR	228

Search completed: June 15, 2005, 16:13:41  
Job time : 159 secs

***This Page Blank (uspro)***



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 15, 2005, 15:50:48 ; Search time 42 Seconds  
(without alignments)  
545.649 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALGLOKVRNTEBLRKEKSR.....TESSLPSWLMALNRKNCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backlist1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	498	30.4	827	4	US-09-919-039-149
2	495	30.3	245	4	US-09-438-833-3
3	495	30.3	330	4	US-09-438-833-4
4	495	30.3	623	4	US-09-967-388-2
5	495	30.3	652	4	US-09-438-833-5
6	495	30.3	813	4	US-09-438-833-12
7	495	30.3	826	1	US-08-785-241-6
8	495	30.3	826	2	US-08-480-473B-2
9	495	30.3	826	3	US-08-915-213-2
10	495	30.3	826	3	US-09-148-547-2
11	495	30.3	826	3	US-09-235-217-2
12	495	30.3	826	3	US-09-380-662-23
13	495	30.3	826	4	US-09-438-833-1
14	495	30.3	826	4	US-09-702-705-330
15	495	30.3	826	4	US-09-736-457-330
16	495	30.3	826	4	US-09-383-581-2
17	495	30.3	826	4	US-09-614-124B-330
18	495	30.3	826	4	US-09-671-325-330
19	495	30.3	826	4	US-09-589-184-330
20	495	30.3	826	4	US-09-658-824-330
21	495	30.3	826	4	US-09-959-873B-18
22	495	30.3	826	4	US-09-949-016-6089
23	495	30.3	826	4	US-09-967-388-4
24	495	30.3	826	5	PCT-US96-10251-2
25	489	29.9	810	1	US-08-785-241-7
26	477.5	28.5	875	1	US-08-785-241-5
27	466	28.5	485	3	US-09-374-454-4

28	466	28.5	870	1	US-08-785-241-4	Sequence 4, Appli
29	466	28.5	370	2	US-09-374-454-6	Sequence 6, Appli
30	444	27.1	373	3	US-08-480-473B-3	Sequence 3, Appli
31	444	27.1	373	3	US-08-915-213-3	Sequence 3, Appli
32	444	27.1	373	3	US-09-235-217-3	Sequence 3, Appli
33	444	27.1	373	5	PCT-US96-10251-3	Sequence 3, Appli
34	444	27.1	805	2	US-08-480-473B-4	Sequence 4, Appli
35	444	27.1	805	3	US-08-915-213-4	Sequence 4, Appli
36	444	27.1	805	3	US-09-235-217-4	Sequence 4, Appli
37	444	27.1	805	5	PCT-US96-10251-4	Sequence 4, Appli
38	345.5	21.1	613	4	US-09-438-833-6	Sequence 6, Appli
39	323	19.7	756	4	US-09-438-833-11	Sequence 11, Appli
40	297	18.2	595	4	US-09-949-016-7205	Sequence 7205, Ap
41	294	18.0	590	2	US-08-785-310A-5	Sequence 5, Appli
42	293.5	17.9	594	2	US-08-785-310A-6	Sequence 6, Appli
43	281	17.2	1507	4	US-09-914-259-37	Sequence 37, Appli
44	263	16.1	238	4	US-09-949-016-9161	Sequence 9161, Ap
45	256	15.6	570	4	US-09-923-684-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-919-039-149  
; Sequence 149, Application US/09919039  
; Patent No. 6727066  
; GENERAL INFORMATION:  
; APPLICANT: Kaefer, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919, 039  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 149  
; LENGTH: 827  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Incyte ID No. 6727066 1250434CD1  
US-09-919-039-149

Query Match 30.4%; Score 498; DB 4; Length 827;  
Best Local Similarity 46.2%; Pred. No. 1.4e-45;  
Matches 115; Conservative 26; Mismatches 60; Indels 48; Gaps 6;  
QY 9 RSNTELRKESRDARSRSQETEVLYOLAHTLPFARGVAHLDKASIMRLTTSYLRMR 68  
13 RISRRKESKSRDAARSRSKESSEVFYELAHQLPLPHNVSSHLDKASVMSLTTSYLRVRK 72  
Db 69 LCAAGKRGKRGATRLPEGPGFRHGTGRGRGLPVKCOA-----PGPOSVD 118  
73 LIDAG-----DLIDBDMKQNMCFYLKALDGFVMVL 104  
QY 119 LCSSSLIH-----NPTPG-TNFSLELIGHSIPFTHPCDOEELQDALTPPNTSKKLEA 172  
Db 105 TDDGDIMYISDNNVKWGLTQF--ELTGHSEVPFTHPCDHEBRREMLTHNGGLVKKGEQ 162  
QY 173 PTERHPSLRKSTLTSGRTLNKAAITWVKLHSGHMRAYKPPAQTSPPGSPSPLOC 232  
Db 163 NTPRSFPLRKCTLTSGRTNKSATWVKLHCTGHIVDTNSNPQCG--YKKPMTIC 220  
QY 233 LVTLICEAIP 241  
Db 221 LVTLICEPI 229  
RESULT 2  
US-09-438-833-3

```

? Sequence 3 Application US/09438833
? Patent No. 6436654
? GENERAL INFORMATION:
? APPLICANT: Pharmacia & Upjohn
? TITLE OF INVENTION: Protein variants
? FILE REFERENCE: 1848
? CURRENT APPLICATION NUMBER: US/09/438.833
? CURRENT FILING DATE: 1999-11-12
? NUMBER OF SEQ ID NOS: 15
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 3
? LENGTH: 245
? TYPE: PRT
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: Subdomain
? OTHER INFORMATION: 1-245 of human HIF-1 alpha
? US-09-438-833-3

```

[illegible]

```

RESULT 3
US-09-438-833-4
; Sequence 4, Application US/09438833
; Patent No. 6436654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438.833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Subdomain
; US-09-438-833-4

```

```

QY 71 AACGKKGRATGRLPEGPGCFGRHTRGRGRLPYGKQQA-----PGQSVTLG 120
    ||| |||
Db 74 DAG-----DLIDEDMKRQNNCFYLKALDGFVWVLT 105

QY 121 SSSLIR-----NPTQG-TNFSLELIGHSIFDFIHPCDQELQDALTRPNLSKKLEAPT 174
    ||| |||
Db 106 DGDMIYIDNNVKKWGLNQF--ELTGHVSVPDFTHPCDHEKREWLTTRNGVLVKKGKEQNT 163

QY 175 ERFSLIRMSSTLTSGRGTNLKAAWVKTLHSGGMRKRAYKPPAQSTSPGSRSEPPDCLV 234
    ||| |||
Db 164 QRSFFLRMCSTLTSGRTNINIKSATWVKVLTGHIHYVDITNSNQPGG--YKKEPMTCIV 221

QY 235 LICEAP 241
    ||| |||
Db 222 LICEPIP 228

```

```

: RESULT 4
: US-09-967-388-2
: Sequence 2, Application US/09967388
: Patent No. 6838430
:
: GENERAL INFORMATION:
:
: APPLICANT: JEFFEREY M. ARBETT
:
: TITLE OF INVENTION: USE OF HIF-1ALPHA VARIANTS TO ACCELERATE
: FILE REFERENCE: UC077.001A
:
: CURRENT APPLICATION NUMBER: US/09/967.388
:
: CURRENT FILING DATE: 2001-09-28
:
: NUMBER OF SEQ ID NOS: 4
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 2
:
: LENGTH: 623
:
: TYPE: PRT
:
: ORGANISM: HUMAN
:
: US-09-967-388-2

```

	Query Match	30.3%;	Score	495;	DB	4;	Length	623;		
	Best Local Similarity	46.2%;	Pred.	No. 2e-45;						
	Matches	114;	Conservative	26;	Mismatches	59;	Indels	48;	Gaps	6
Oy	11	NPELRKESRDAARRSROETEVLYQLAHTLPFARGVSAHLDDKASIMRLTISTYLRMHRIC	70							
	:	:::::	:	:	:	:	:	:	:	:
Db	14	SSRRRKESKRDAARRRSRESEVFTEIHAOLPLPHNVSSHLDDKASVMRLTISTYLRYVKLL	73							
Oy	71	AAGCKRGATGRTLLPEGCGFRHGHTRGRGHGLPVGKCOQA-----PGQSVLDLC	120							
	:	:	:	:	:	:	:	:	:	:
Db	74	DAG-----DLDIEDMDKAYNNCFYALKALDFWVLTD	105							
Oy	121	SSSLIH----NPTQG-TNFSLLELIGHSIFDPFHPCDEELQDALITRPMLSKKLEAPT	174							
	:	:	:	:	:	:	:	:	:	:
Db	106	DGMVIYISDNVRKYWGLOTGF--ELIGHSVFPFTPRCHEEREMLTIRNGILVKKGEQNT	163							
Oy	175	ERRFSLRMSKSTLYTSRGRTLTNLKAATWKVLAHCSGMRAYKPPACSPSGSRPSEPLOCTV	234							
	:	:	:	:	:	:	:	:	:	:
Db	164	QRSFFPLRMKCLTITSGRMTNIKSATWVKVLIHQTHIHYYDTNSNQPGGG--YKKEPMTCIV	221							
Oy	235	LICEATP	241							
	:	:	:	:	:	:	:	:	:	:
Db	222	LICEPTP	228							

```

RESULT 5
US-09-438-833-5
; Sequence 5, Application US/09438833
; Patent No. 6438654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438,833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1

```





Best Local Similarity 46.2%; Pred. No. 3e-45;

3  
1  
0  
4  
3  
2  
5

Best Local Similarity 46.2%; Pred. No. 3e-45;

```
US-09-438-833-1
; Sequence 1, Application US/09438833
; Patent No. 6436654
; GENERAL INFORMATION:
; APPLICANT: Pharmacia & Upjohn
; TITLE OF INVENTION: Protein variants
; FILE REFERENCE: 1848
; CURRENT APPLICATION NUMBER: US/09/438,833
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; VOLUME: 92
; PAGES: 5510-5514
; DATABASE ACCESSION NUMBER: GenBank U22431
; DATABASE ENTRY DATE: 1995-06-28
; US-09-438-833-1
```

```
Query Match      30.3%; Score 495; DB 4; Length 826;
Best Local Similarity 46.2%; Pred. No. 3e-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
QY 11 NTELRKESRDARSRSQTEVLYOLAHLPARGVSAHLDKASIMRLTISYLRMRLC 70
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERRKESRDARSRSKSEVFEYELAHQLPPLHNVSSHLDKASVRLTISYLRVKKL 73
QY 71 AAGKRGATGRLLPBGPGFRHGRHGLPVGKCOQA-----PGQSVDLG 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLDIEDMKAKQNNCFYLKALDGFVWVLTLD 105
QY 121 SSSLIH-----NPTPG-TNFSLEIGHSTDFTHPCDQELDQALTPRPNLSTKKLEAPT 174
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKMGITQF--ELTGHSVDFTHPCDHEEREMLTIRNGLVKKGEQNT 163
QY 175 ERHPSLRMKSITLSRGRTLNKAAATWKVLCSGHMAAYKPPAQTSPAGSRSEPPLOCLV 234
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 QRSFPLRMKCTLSRGRTNMIKSAATWKVLCGTHIHVDTNSNQPCG--YKKRPMTCLV 221
QY 235 LICEAIP 241
   |||||
Db 222 LICEPIP 228
```

```
RESULT 14
US-09-702-705-330
; Sequence 330, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedavick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: PatSeq for Windows Version 3.0
; SEQ ID NO 330
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
```

```
US-09-702-705-330
Query Match      30.3%; Score 495; DB 4; Length 826;
Best Local Similarity 46.2%; Pred. No. 3e-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
QY 11 NTELRKESRDARSRSQTEVLYOLAHLPARGVSAHLDKASIMRLTISYLRMRLC 70
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERRKESRDARSRSKSEVFEYELAHQLPPLHNVSSHLDKASVRLTISYLRVKKL 73
QY 71 AAGKRGATGRLLPBGPGFRHGRHGLPVGKCOQA-----PGQSVDLG 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLDIEDMKAKQNNCFYLKALDGFVWVLTLD 105
QY 121 SSSLIH-----NPTPG-TNFSLEIGHSTDFTHPCDQELDQALTPRPNLSTKKLEAPT 174
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKMGITQF--ELTGHSVDFTHPCDHEEREMLTIRNGLVKKGEQNT 163
QY 175 ERHPSLRMKSITLSRGRTLNKAAATWKVLCSGHMAAYKPPAQTSPAGSRSEPPLOCLV 234
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 QRSFPLRMKCTLSRGRTNMIKSAATWKVLCGTHIHVDTNSNQPCG--YKKRPMTCLV 221
QY 235 LICEAIP 241
   |||||
Db 222 LICEPIP 228
```

```
RESULT 15
US-09-736-457-330
; Sequence 330, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedavick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: PatSeq for Windows Version 3.0
; SEQ ID NO 330
; LENGTH: 826
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-736-457-330
```

```
Query Match      30.3%; Score 495; DB 4; Length 826;
Best Local Similarity 46.2%; Pred. No. 3e-45;
Matches 114; Conservative 26; Mismatches 59; Indels 48; Gaps 6;
```

```
QY 11 NTELRKESRDARSRSQTEVLYOLAHLPARGVSAHLDKASIMRLTISYLRMRLC 70
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 SSERRKESRDARSRSKSEVFEYELAHQLPPLHNVSSHLDKASVRLTISYLRVKKL 73
QY 71 AAGKRGATGRLLPBGPGFRHGRHGLPVGKCOQA-----PGQSVDLG 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 74 DAG-----DLDIEDMKAKQNNCFYLKALDGFVWVLTLD 105
QY 121 SSSLIH-----NPTPG-TNFSLEIGHSTDFTHPCDQELDQALTPRPNLSTKKLEAPT 174
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 106 DGDMIYISDNVNMKMGITQF--ELTGHSVDFTHPCDHEEREMLTIRNGLVKKGEQNT 163
QY 175 ERHPSLRMKSITLSRGRTLNKAAATWKVLCSGHMAAYKPPAQTSPAGSRSEPPLOCLV 234
   :|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

Db 164 QRSFPLMKCTLTSRGRTMNKSAWIKVLCCTGHIHYVDINSNOPCG--YKPPMTCLV 221  
Qy 235 LICEAIP 241  
Db 222 LICEPIP 228

Search completed: June 15, 2005, 16:01:42  
Job time : 43 secs

This Page Blank (copy)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 15, 2005, 15:36:58 ; Search time 162 Seconds  
(without alignments)  
732.935 Million cell updates/sec

Title: US-09-896-791B-3

Perfect score: 1636  
Sequence: 1 MALGLQRVRSNTBLRKEKSR.....TESSLPSWVLMALNRKNCPG 307

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1636	100.0	307	5	AAU75902 Mouse inh
2	1636	100.0	307	6	ABR62227 Mouse inh
3	1278	78.1	324	6	ABR83343 Human NOV
4	855.5	52.3	662	2	AAV06295 Mouse tra
5	831	50.8	632	4	AAH93326 Human pro
6	803	49.1	790	6	AAH93326 Human pro
7	802	49.0	667	5	AAE24222 Human nuc
8	797.5	48.7	747	6	AAO16417 Human HIF
9	797	48.7	407	5	ABG66737 Human nov
10	795	48.6	705	6	AAE34826 Human hyp
11	573	35.0	648	4	AAH94934 Human pro
12	518.5	31.7	199	4	AAH93710 Human pro
13	498	30.4	513	8	AD116229 Human nuc
14	498	30.4	827	7	ADE25733 Human pro
15	498	30.4	827	8	ADE76984 Human pro
16	497	30.4	825	7	ADDD4855 Rat Prote
17	495	30.3	245	7	AAV94627 HIF-1alp
18	495	30.3	330	3	AAV94628 HIF-1alp
19	495	30.3	623	6	ABR40379 Human HIF
20	495	30.3	642	3	AAV84168 A variant
21	495	30.3	652	3	AAV94629 HIF-1alp
22	495	30.3	669	3	AAV84167 A variant
23	495	30.3	697	3	AAV84166 A variant
24	495	30.3	701	3	AAV84173 A variant
25	495	30.3	710	3	AAV84172 A variant

26	495	30.3	724	3	AAV84171 A variant
27	495	30.3	735	6	ABR82375 Hypoxia-1
28	495	30.3	735	6	ADN75066 Human hyp
29	495	30.3	749	3	AAV84170 A variant
30	495	30.3	789	3	AAV84169 A variant
31	495	30.3	789	6	ADAI8535 Human hyp
32	495	30.3	789	6	ADAI8533 Human hyp
33	495	30.3	789	6	ADAI8534 Human hyp
34	495	30.3	813	3	AAV94636 HIF-1alp
35	495	30.3	826	2	AAV06557 Human hyp
36	495	30.3	826	2	AAH80418 amino aci
37	495	30.3	826	2	AAV06289 Human tra
38	495	30.3	826	2	AAV06289 A wild ty
39	495	30.3	826	3	AAV94640 Human hyp
40	495	30.3	826	4	AAV76854 Human lun
41	495	30.3	826	5	AAU85509 Clone #19
42	495	30.3	826	5	AAU77620 Human hyp
43	495	30.3	826	5	AAU77602 Human hyp
44	495	30.3	826	5	AAU77619 Human hyp
45	495	30.3	826	5	AAU77607 Human hyp

#### ALIGNMENTS

RESULT 1	AAU75902 standard; protein; 307 AA.
ID	AAU75902;
AC	AAU75902;
DT	08-MAY-2002 (first entry)
XX	Mouse inhibitory PAS domain (IPAS) protein.
DE	Inhibitory PAS domain protein; IPAS; cardiant; vasotropic;
XX	cardiovascular; cerebroprotective; ophthalmological; HIF-1alpha;
KW	hypoxia-induced factor 1alpha; vascular endothelial growth factor;
KW	angiogenesis; ischemic cardiovascular lesion; stroke;
KW	diabetic microvascular disease; tumour; mouse.
XX	Mus musculus.
OS	WO200202609-A1.
XX	10-JAN-2002.
PN	19-JUN-2001; 2001WO-SB001387.
XX	06-JUL-2000; 2000SE-00002551.
PR	(BIOV-) BIOVITRUM AB.
PA	Berkenstam A, Bertilsson G, Poellinger L;
PI	WPI; 2002-164523/21.
XX	N-PSDB; ABK14502.
DR	New nucleic acid encoding inhibitory PAS domain protein, useful for
XX	identifying specific inhibitors for treating e.g. angiogenesis or tumor
PT	growth.
XX	Claim 3; Fig 1; 4app; English.
PS	The invention describes an isolated nucleic acid encoding the
XX	biologically active inhibitory PAS domain protein or its functionally
CC	equivalent modifications. IPAS forms a non-functional heterodimeric
CC	complex with HIF-1alpha (hypoxia-induced factor 1alpha), impairing
CC	interaction between HIF-1alpha and hypoxia-response elements in genes,
CC	e.g. the gene for vascular endothelial growth factor, so contributes to
CC	control of hypoxic signalling. The nucleic acid and its encoded
CC	polypeptides, are used to identify agents that activate expression of the
CC	gene or stimulate activity of the protein. These agents are useful for

CC inhibiting angiogenesis, particularly where associated with ischaemic  
CC cardiovascular lesions, stroke or diabetic microvascular diseases, and  
CC tumour growth. This is the amino acid sequence of the mouse inhibitory  
CC PAS domain protein (IPAS), described in the method of the invention  
XX  
SQ Sequence 307 AA;

Query Match	100.0%;	Score 1636;	DB 5,	Length 307;
Best Local Similarity	100.0%;	Pred. No. 1.9e-160;		
Matches 307; Conservative	0;	Mismatches	0;	Gaps 0;

Qy	1	MALGLQVRVSENTELREKESRDPAARSRSQETEVLLQLAHNTLPFARGVSAHLIDKXSIINRLT	60
Db	1	MALGLQVRVSENTELREKESRDPAARSRSQETEVLLQLAHNTLPFARGVSAHLIDKXSIINRLT	60
Qy	61	ISYLRMRLLCAAGKRGKGRATGRLLPEPGSGFPHGTHRRGRHGLPVKCKQQAPEGQSVDLG	120
Db	61	ISYLRMRRLCAAGKRGKGRATGRLLPEPGSGFPHGTHRRGRHGLPVKCKQQAPEGQSVDLG	120
Qy	121	SSSLINHPPTGNTFSELIGHSIFPDIHPCDDDEILQDLATLRPNLSKKGLEAPERRHPSL	180
Db	121	SSSLINHPPTGNTFSELIGHSIFDPDIHPCDDDEILQDLATLRPNLSKKGLEAPERRHPSL	180
Qy	181	RMSKSTLSRGRITNLKAATWKLHCSGHRMAYKPAQOTSPLGSPSPSEPPLOCLVLTCEAI	240
Db	181	RMSKSTLSRGRITNLKAATWKLHCSGHRMAYKPAQOTSPLGSPSPSEPPLOCLVLTCEAI	240
Qy	241	POKPFHDGATLGLPOEKTPISTLFTPLMAYALLCLVKRPVPVOVLQCKGESSLSPSMTLMAL	300
Db	241	POKPFHDGATLGLPOEKTPISTLFTPLMAYALLCLVKRPVPVOVLQCKGESSLSPSMTLMAL	300
Qy	301	NRKNCPSG 307	
Db	301	NRKNCPSG 307	

RESULT 2  
ABR62227  
ID ABR62227 standard; protein; 307 AA.

AC	ABR62227;
XX	
DT	08-SEP-2003 (first entry)

DE Mouse inhibitory PAS (Per/Arnt/Sim) domain protein IPAS.

KM Mouse; IPAS; inhibitory PAS domain protein; hypoxia; vasotropic;  
KM cardant; vulnerary; cerebroprotective; antiulcer; gene therapy.

Mus sp.

FH	Key	Location/Qualifiers
----	-----	---------------------

```

44 region 17.125
FT /label= Basic

```

FT	20. : 41	/label= Helix
F1	region	

F1	Region	42.33
FT	/label= Loop	

```

FI Region 56. : 6 /
FT /label= Helix

```

FT Domain 103. :170 /label= PAS A

FT	Domain	235. .307
FT		/label = PAS B

XX  
PN W02003045440-A1

XX  
PD 05-JTN-3003

XX  
XX  
29-NOV-2003 12:00:00

xx  
c c c c c c c c c c c c c c c c

XX

PA (ANGI-) ANGIOGENETICS SWEDEN AB  
XX  
XX PI Makino Y, Cao Y, Poellinger L,  
XX  
DR WPI; 2003-505168/47.  
DR N-PSDB; ACCC3867.  
DR

PT New pharmaceutical composition comprising IP<sub>2</sub>S, useful for increasing  
PT angiogenesis, or for treating a hypoxia-mediated condition in a cell,  
PT group of cells or organisms under hypoxic conditions, e.g. stroke or  
PT coronary heart disease.

PS Claim 4; Fig 1; 45pp; English

The present sequence of a novel murine basic helix-loop-helix/PAS (Per/Arnt/Sim) protein, termed inhibitory PAS domain protein or IPAS, that is related to hypoxia-inducible transcription factors (HIFs). IPAS has no transactivation function and demonstrates dominant negative regulation of HIF-mediated control of gene expression. Ecotypic expression of IPAS in hepatoma cells selectively impaired induction of genes involved in adaptation to a hypoxic environment, and also resulted in retarded tumour growth and reduced tumour vascular density *in vivo*. In mice, IPAS was predominantly expressed in Purkinje cells and in corneal epithelium. Application of an IPAS antisense oligonucleotide to the mouse cornea induced angiogenesis under normoxic conditions, and unmasked hypoxia-dependent induction of vascular endothelial growth factor gene expression in hypoxic cornea cells, indicating a novel regulation of angiogenesis and maintenance of an avascular phenotype. Claimed pharmaceutical compositions comprising IPAS antisense sequences can be used to treat increased angiogenesis, to stimulate HIF- $\alpha$  function, to treat hypoxia-related conditions such as ischaemia, coronary heart disease, wound healing, stroke or diabetic ulceration, and to maintain normal cell functions under hypoxia

Sequence 307 AA;

Query Match	100.0%	Score 1636;	DB 6	Length 307;
Best Local Similarity	100.0%	Pred. No. 1.9e-16;		
Matches 307; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0

[illegible]

61 ISYLRMRLCAAGCKRGATGRLPEGPGGFRHGTNRGRHGLPVCKCQAPGQSVDC 120

Db 61 ISYLRMRLCAAGKRGRTGRLLPEGPGGFRHGTHRRGRHGLPVGKCQDAPGPQSVDL C 120

OV 121 SSSLHNPTPGTNFSLIGHSIFDFIHPCDQEELODALTPRPNLSSKKLEAFTERHFSL 180

D<sub>0</sub> 121 SSSLIHNPTPGTNFSLIELIGHSIFDFIHPDOEELODALTPRNLSKKLEAPTERHESL 180

181 RMKSTLSRGRTLNKATWKLHCSGMBRAYKPPAOTSPAGSPRSEPLCLVLC EAI 240

Db 181 RMKSTLT SRGRTLNKATWVLIHCSGMRAYKPPAOTSPAGSPRSEPILOCVLI CEAT 240

241 ПОРЕНДАТИГИ.ПОРЕКЪРЪТИ.ЕТБИ.КРАЛ.Ц.УКРМДВОУТ.ОКГГТЕСС.ПСУУТ.УАТ. 300

341 ПО РЕШЕНИИ ПОКРЕПИТЕ ЕДИН КРАЙ С КРЕПЕЖНЫМИ ГРЕЙДЕРИ. ПОВТОРИТЕ. 300

301 NBYKCBQ 307

**Figure 1**

### RESULT 3

ID	ABR83343	standard; protein; 324 AA.
----	----------	----------------------------

AC ABR83343;

DT 06-OCT-2003 (first entry)

DE Human NOV9a protein SEQ ID NO:20.  
 XX  
 XX Human; NOVX; cytosolic; antidiabetic; neuroprotective; antiparkinsonian;  
 KM anorectic; gene therapy; vaccine; cancer; neurodegenerative disorder;  
 KM Parkinson's disease; metabolic disorder; diabetes; obesity;  
 KM tissue typing.  
 OS Homo sapiens.  
 XX  
 XX WO2003052061-A2.  
 PN  
 XX  
 PD 26-JUN-2003.  
 XX  
 XX 03-DEC-2002; 2002WO-US038821.  
 PF  
 XX 17-DEC-2001; 2001US-0341477P.  
 PR 17-DEC-2001; 2001US-0341540P.  
 PR 20-DEC-2001; 2001US-0342582P.  
 PR 31-DEC-2001; 2001US-0344903P.  
 PR 17-APR-2002; 2002US-0373288P.  
 PR 15-MAY-2002; 2002US-0380981P.  
 PR 17-MAY-2002; 2002US-0381495P.  
 PR 28-MAY-2002; 2002US-0383744P.  
 PR 29-MAY-2002; 2002US-0384024P.  
 PR 07-AUG-2002; 2002US-0401788P.  
 PR 26-AUG-2002; 2002US-0406353P.  
 PR 31-OCT-2002; 2002US-0422756P.  
 PR 02-DEC-2002; 2002US-00307928.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 PA  
 XX  
 PI Alsobrook JP, Anderson DW, Boldog FL, Burgess CE, Catterton E;  
 PI Edinger SR, Gorman L, Guo X, Ji W, Kekuda R, Li L, Paturajan M;  
 PI Rieger DK, Shenoy SG, Spytek KA, Vernet CAM, Voss EZ, Zhong M;  
 DR WPI, 2003-533005/50.  
 DR N-PSDB; ACP06242.  
 XX  
 XX New NOVX polypeptide, useful for preparing a composition for treating or  
 PT preventing e.g. cancer, neurodegenerative disorders such as Parkinson's  
 PT disease, or metabolic disorders such as diabetes or obesity, or for  
 PT tissue typing.  
 XX  
 XX  
 PS Claim 1; Page 130; 190pp; English.  
 XX  
 XX ACP06242 encode the human NOVX proteins given in ABR83334 to  
 CC ABR83343, designated NOV1a, NOV2a, NOV3a, NOV4a, NOV5a, NOV6a,  
 CC NOV7a, NOV8a and NOV9a respectively. NOVX sequences can have cytosolic,  
 CC antidiabetic, neuroprotective, antiparkinsonian and anorectic activities,  
 CC and can be used in vaccines and gene therapy. The NOVX polypeptides can  
 CC be used for preparing a composition for treating or preventing a  
 CC pathology associated with the NOVX-polypeptides e.g. cancer,  
 CC neurodegenerative disorders such as Parkinson's disease, or metabolic  
 CC disorders such as diabetes or obesity, or for tissue typing  
 CC  
 XX  
 XX Sequence 324 AA;  
 SQ  
 Query Match 78.1%; Score 1278; DB 6; Length 324;  
 Best Local Similarity 79.9%; Pred. No. 2.5e-123;  
 Matches 250; Conservative 14; Mismatches 41; Indels 8; Gaps 3;  
 QY 1 MALGLQVRNSELTKKESRDAAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 DB 1 MALGLQVRNSELTKKESRDAAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLT 60  
 QY 61 ISYLRMRLCAAGGKRGATGRLLPBGPGFRGCTHRGRHGLPVGKCOQAPGQSYDLC 120  
 DB 61 ISYLRMRLCAAGGKRGATGRLLPBGPGFRGCTHRGRHGLPVGKCOQAPGQSYDLC 120  
 QY 121 SSSLINHTPGTNSLGLGHSIFDPIHPCDQELQDALTPRPVLSSKKKL EAPTERHFSL 180  
 DB 121 SSSLINHTPGTNSLGLGHSIFDPIHPCDQELQDALTPRPVLSSKKKL EAPTERHFSL 180  
 121 SPSLPRIPTGINSLELIGHISIFDFIHPCDQELQDALTPOALSRKKEAFTERCSFL 180

QY 181 RMKSTLTSRGRTINLKAATKVLHCSGHMRAVYPAPQTSPPAGSPRSEPPQLCVLICEAT 240  
 DB 181 RMKSTLTSRGRTINLKAATKVLHCSGHMRAVYPAPQTSPPAGSPRSEPPQLCVLICEAT 240  
 QY 241 POLPFDHGAFLGAPQETPISTFTPLMKL-----LCVKKRPVQY-LOGKTESLPS 294  
 DB 241 PSLSTDGATLGPQEA--SPLSPTYPVCGRHSCLCFVREWPDIQAGMGSDPSLLPL 298  
 QY 295 WVLWALNRKNCPG 307  
 DB 299 EVMWALNGNCAG 311  
 RESULT 4  
 AAY06295  
 ID AAY06295 standard; protein; 662 AA.  
 XX  
 AC AAY06295;  
 XX  
 DT 23-AUG-1999 (first entry)  
 XX  
 DE Mouse transcription regulator MOP7.  
 XX  
 KM MOP7, member of the PAS superfamily; bHLH-PAS; mouse;  
 KM transcription regulator; hypoxia inducible factor 3 alpha.  
 XX  
 OS Mus musculus.  
 XX  
 PN W09928464-A2.  
 XX  
 PD 10-JUN-1999.  
 XX  
 PF 27-NOV-1998; 98MO-US025314.  
 XX  
 PR 28-NOV-1997; 97US-0066863P.  
 XX  
 PA (WISC) WISCONSIN ALUMNI RES FOUND.  
 XX  
 PI Bradford CA, Gu YZ, Hogenesch JB;  
 XX  
 DR WPI, 1999-371120/31.  
 DR N-PSDB; AAX58986.  
 XX  
 PT Developmental signal transduction associated proteins.  
 XX  
 PS Claim 6; Page 101; 106pp; English.  
 XX  
 XX The present sequence represents mouse MOP7, a novel member of the PAS  
 CC superfamily, where PAS stands for PER/ARNT/SIM domains. MOP7 cDNA (see  
 CC AAX58986) was identified in a search of murine ESTs designed to identify  
 CC bHLH-PAS proteins, and by PCR amplification of lung cDNA. MOP7 was  
 CC characterised as hypoxia inducible factor 3 alpha (HIF 3 alpha). Its  
 CC expression profile is distinct from that of HIF 1 alpha (see AAY06289),  
 CC HIF 2 alpha (see AAY06290), MOP3 (see AAY06291), Ah receptor and Ah  
 CC receptor nuclear translocator (ARNT), suggesting a different functional  
 CC role. MOP7 probably regulates the same genes as HIF 1 alpha and 2 alpha,  
 CC as evidenced by its dimerisation with the same partners (ARNT, MOP3) and  
 CC recognition of the same core response element. MOP7 may have a functional  
 CC role associated with response to low oxygen in the tissues in which it is  
 CC expressed. The invention provides novel MOPs 2-9 nucleic acids (see  
 CC AAX58981-88) and proteins (see AAY06289-97). These are useful in a  
 CC variety of research, diagnostic and therapeutic applications. Several of  
 CC the MOPs are alpha-class hypoxia-inducible factors. Others are involved  
 CC in circadian signal transduction  
 XX  
 XX Sequence 662 AA;  
 SQ  
 Query Match 52.3%; Score 855.5; DB 2; Length 662;  
 Best Local Similarity 68.4%; Pred. No. 4.2e-79;  
 Matches 182; Conservative 5; Mismatches 24; Indels 55; Gaps 4;  
 QY 9 RSTELRKESRDAAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMR 68  
 9 RSTELRKESRDAAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLRMR 68

Db 7 RSNTLEKESRDARSRSOETEVLYQLAHTLPFANGVSAHLDKASIMRLTISYLEMR 66  
 Qy 69 LCAAGKRGKRGATGRLLEPGGPFHGTTRGRHGLPVGKCCQAPGPOVDLC----- 120  
 Db 67 LCAAGEMNN-----QVEKGSEPLDACVYKALEGF 94  
 Qy 121 -----SSSLIHNPTPGTNFS-LELIGHSTDFPHPCDQELQDALTPRPVLSKKL 170  
 Db 95 VMVLTABGDMAVYLEENYSKILGLSQLELIGHSTDFPHPCDQELQDALTPRPVLSKKL 154  
 Qy 171 EAPTERHPSLRMKSSTLTSRGRTLNKKAATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSEPL 230  
 Db 155 EAPTERHPSLRMKSSTLTSRGRTLNKKAATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSEPL 214  
 Qy 231 QCLVLICEAIP-----QLPFHDGATL 251  
 Db 215 QCLVLICEAIPHPASLEPPLGRGAFL 240  
 Db  
 RESULT 5  
 AAB93326  
 ID AAB93326 standard; protein; 632 AA.  
 XX  
 AC AAB93326;  
 XX  
 DT 26-JUN-2001 (first entry)  
 DE Human protein sequence SEQ ID NO:12422.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EPI074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-00116126.  
 XX  
 PR 29-JUL-1999; 99JP-00248036.  
 XX  
 PR 27-AUG-1999; 99JP-00300253.  
 PR 11-JAN-2000; 2000JP-00118776.  
 PR 02-MAY-2000; 2000JP-00183767.  
 PR 09-JUN-2000; 2000JP-00241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ora T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI; 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.  
 XX  
 PS Claim 8; SEQ ID NO 12422; 2537bp + Sequence Listing; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides.

CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 XX  
 SQ Sequence 632 AA;  
 Query Match 50.8%; Score 831; DB 4; Length 632;  
 Best Local Similarity 68.2%; Pred. No. 1.4e-76;  
 Matches 178; Conservative 11; Mismatches 30; Indels 42; Gaps 5;  
 Qy 1 MALGLRVSNTELEKESRDARSRSOETEVLYQLAHTLPFANGVSAHLDKASIMRLT 60  
 Db 1 MALGLRVSNTELEKESRDARSRSOETEVLYQLAHTLPFANGVSAHLDKASIMRLT 60  
 Qy 61 ISYLRMRLCAAGKRGKRGATGRLLEPGGPFHGTTRGRHGLPVGKQ----- 109  
 Db 61 ISYLRMRLCAAG-----EMNQVAGGSEPLDACVYKALEGFVMVL 100  
 Qy 110 QARGPOVDLCSSSLIHNPTPGTNFS-LELIGHSTDFPHPCDQELQDALTPRPVLSKK 168  
 Db 101 TARGDMAY-----LSENVSKILGLSQLELIGHSTDFPHPCDQELQDALTPQQLSR 154  
 Qy 169 KLEAPTERHPSLRMKSSTLTSRGRTLNKKAATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSE 228  
 Db 155 KLEAPTERHPSLRMKSSTLTSRGRTLNKKAATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSE 214  
 Qy 229 PLOCLVLICEAIPQLPFHDGATL 249  
 Db 215 PLOCLVLICEAIP-----HPGS 231  
 Db  
 RESULT 6  
 AA016439  
 ID AA016439 standard; protein; 790 AA.  
 XX  
 AC AA016439;  
 XX  
 DT 10-APR-2003 (first entry)  
 DE Human nucleic acid-associated protein (NAAP) - SEQ ID NO 36.  
 XX  
 KW Human; nucleic acid-associated protein; NAAP; arteriosclerosis;  
 KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
 KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
 KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;  
 KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
 KW Crohn's disease; transgenic animal; animal model.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W02003000864-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 20-JUN-2002; 2002WO-US021179.  
 XX  
 PR 22-JUN-2001; 2001US-0300518P.  
 PR 29-JUN-2001; 2001US-0301787P.  
 PR 29-JUN-2001; 2001US-0301792P.  
 PR 29-JUN-2001; 2001US-0301892P.  
 PR 29-JUN-2001; 2001US-0301893P.  
 PR 06-JUL-2001; 2001US-0303405P.  
 PR 06-JUL-2001; 2001US-0303442P.  
 PR 15-MAR-2002; 2002US-0364438P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;



XX AA016417;  
AC 10-APR-2003 (first entry)  
DT  
XX  
XX Human nucleic acid-associated protein (NAAP) - SEQ ID NO 14.  
DE  
XX  
XX Human, nucleic acid-associated protein; NAAP; arteriosclerosis;  
KM cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;  
KM cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;  
KM mental retardation; neurological disorder; Alzheimer's disease; epilepsy;  
KM Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;  
KM Crohn's disease; transgenic animal; animal model.  
XX  
OS Homo sapiens.  
XX  
XX WO2003000864-A2.  
PN  
XX  
XX 03-JAN-2003.  
PD  
XX  
XX 20-JUN-2002; 2002WO-US021179.  
PF  
XX  
XX 22-JUN-2001; 2001US-0300518P.  
PR 29-JUN-2001; 2001US-0301787P.  
PR 29-JUN-2001; 2001US-0301792P.  
PR 29-JUN-2001; 2001US-0301892P.  
PR 29-JUN-2001; 2001US-0301893P.  
PR 06-JUL-2001; 2001US-0303405P.  
PR 06-JUL-2001; 2001US-0303442P.  
PR 15-MAR-2002; 2002US-0364438P.  
XX  
XX (INCYTE GENOMICS INC.  
PA  
PI Gandhi AR, Swarnakar A, Hafalia AUA, Warren BA, Emerling BM,  
PI Ariyazu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IU,  
PI Rankumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn KR;  
PI Borowak ML, Yao MG, Wajia NK, Bandman O, Lal PG, Becha SD, Lee SY,  
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebardjian Y, Lu Y;  
XX  
XX WPI; 2003-201420/19.  
DR N-PSDB; AAL51567.  
XX  
XX New nucleic acid-associated proteins and polynucleotides, useful for  
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),  
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune  
PT disorders (e.g. AIDS).  
XX  
XX  
XX Claim 1; Page 232-234; 312pp; English.  
XX  
XX The invention comprises the amino acid and coding sequences of human  
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of  
CC the invention are useful for diagnosing, treating or preventing disorders  
CC associated with aberrant expression of NAAP, such as: cell proliferative  
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis  
CC or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia  
CC or mental retardation); neurological disorders (e.g. Alzheimer's disease,  
CC Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders  
CC (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of  
CC the invention are useful for creating transgenic animals to model human  
CC disease. The present amino acid sequence represents a human nucleic acid-  
CC associated protein of the invention  
XX  
XX Sequence 747 AA;  
SQ  
Query Match 48.7%; Score 797.5; DB 6; Length 747;  
Best Local Similarity 60.4%; Pred. No. 5.2e-73;  
Matches 180; Conservative 10; Mismatches 35; Indels 73; Gaps 6;  
QY 1 MALGLQVR-----SNTELRK 16  
DB 1 MALGLQVRPALSGVISPPCAPTRNSHPGCTASPPAPGWPFSQSGPGRWSTELRK 60  
QY 17 EKSRDARRSRSSQETVLYQLAHTLPFARGVSAHLDRASIMRLTISTYLRHRLCAAGGR 76

DB 61 EKSRDARRSRSSQETVLYQLAHTLPFARGVSAHLDRASIMRLTISTYLRHRLCAAGGR 120  
QY 77 GRATGR-----LLEPGCGFGRHGRGRGLPVKGCQAPGPOSVDLCSSLIHNPPTGT 132  
DB 121 QVAGENHMMLLKALBGFVMVLTAG-----DM--AIVSENVSKHL 160  
QY 133 NFS-LELIGHSTDFPHPCDEELQDALTPRPNLAKKLEAPTERFSLRMSSTLTSRGR 191  
DB 161 GLSQLELIGHSITDFPHPCDEELQDALTPQOTLSRKVEAPTERCFSLRMSSTLTSRGR 220  
QY 192 TLNLKATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSPPOCLYLICEAIPQLPFDGA 249  
DB 221 TLNLKATWVVLHCSGHRMAYKRPAPQTSPPAGSPRSPPOCLYLICEAIP----HFGS 274  
RESULT 9  
ABG66737  
ID ABG66737 standard; protein; 407 AA.  
AC  
XX  
XX ABG66737;  
DT  
XX  
XX 30-AUG-2002 (first entry)  
DE  
XX  
XX Human novel polypeptide #72.  
XX  
XX Human, inflammatory condition; shock; sepsis; immune response; cancer;  
KM wound healing; central nervous system disease; haematopoiesis;  
KM peripheral nervous system disease; amyotrophic lateral sclerosis; tendon;  
KM myeloid cell disorder; lymphoid cell disorder; platelet disorder; bone;  
KM cartilage; ligament; nerve tissue; ulcer; osteoporosis; osteoarthritis;  
KM bone degenerative disorder; periodontal disease; reperfusion injury;  
KM lung fibrosis; liver fibrosis; autoimmune disorder; bacterial infection;  
KM allergic condition; thrombolysis; thrombosis; coagulation disorder;  
KM fungal infection.  
XX  
XX Homo sapiens.  
XX  
XX WO200244340-A2.  
XX  
XX 06-JUN-2002.  
XX  
XX 30-NOV-2001; 2001WO-US047004.  
XX  
XX  
XX 30-NOV-2000; 2000US-00728952.  
XX  
XX (HYSEQ-) HYSEQ INC.  
XX  
XX Tang YT, Goodrich RW, Liu C, Zhou P, Asundi V, Wang J, Wang D;  
PI Yamazaki V, Ujwal ML, Drmanac RT;  
XX  
XX WPI; 2002-508509/54.  
DR N-PSDB; ABK94961.  
XX  
XX Novel nucleic acids and polypeptides for diagnosis, treatment of  
PT inflammatory, autoimmune, nervous system, myeloid or lymphoid cell  
PT disorders, cancer and promoting wound healing.  
XX  
XX Claim 10; Page 649; 672pp; English.  
XX  
XX The invention relates to human novel polynucleotides and associated  
CC polypeptides. The polynucleotides and polypeptides are useful for  
CC treating inflammatory conditions such as arthritis, nephritis, Crohn's  
CC disease, ischaemia-reperfusion injury, shock, sepsis, immune responses  
CC and cancer and for promoting wound healing. The sequences are used to  
CC induce the proliferation of neural cells and regeneration of nerve and  
CC brain tissue, and are useful for the treatment of central and peripheral  
CC nervous system diseases and neuropathies, such as Alzheimer's disease,  
CC Parkinson's disease, Huntington's disease and amyotrophic lateral  
CC sclerosis. The sequences are involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis, treatment of myeloid or lymphoid  
CC cell disorders and platelet disorders such as thrombocytopenia,  
CC regeneration of bone, cartilage, tendon, ligament and/or nerve tissue

CC growth, tissue repair, healing of burns, incisions, ulcers, treatment of  
 CC osteoporosis, osteoarthritis, bone degenerative disorders and periodontal  
 CC disease. The sequences of the invention are also useful for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, immune deficiencies and disorders  
 CC including severe combined immunodeficiency (SCID), bacterial or fungal  
 CC infections, autoimmune disorders e.g. multiple sclerosis and myasthenia  
 CC gravis, allergic conditions such as asthma, thrombolytic or thrombotic  
 CC and coagulation disorders. Sequences ABG6666-ABG66758 represent human  
 CC novel polypeptides of the invention  
 XX  
 SQ Sequence 407 AA;

Query Match 48.7%; Score 797; DB 5; Length 407;  
 Best Local Similarity 67.9%; Pred. No. 2.4e-73;  
 Matches 171; Conservative 11; Mismatches 28; Indels 42; Gaps 5;

QY 10 SNTLEKREKSRDAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLMHR 69  
 DB 54 STTLEKREKSRDAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLMHR 113  
 QY 70 CAAGKRGATGRLLPEGGGFRHGTNRGRHGLPVGKCQ-----QAPGQSV 118  
 DB 114 CAAAG-----EWNQVAGGGEPLDACYKALBGFVWVLTAEQDMAY- 152  
 QY 119 LCSSSLIHNPTPGTNFS-LELIGHSTFDFIHPDQDELDALTPRPNLSSKKLEAPTERH 177  
 DB 153 -----LSENVSKHLGLSQLELIGHSTFDFIHPDQDELDALTPQQLTSRRKVEAPTERC 207  
 QY 178 FSLRMKSTLTSGRTLNKAAATWVKVLCGSHMRAYKPPAQTSPAGSPREPPLOCLVLI 237  
 DB 208 FSLRMKSTLTSGRTLNKAAATWVKVLCGSHMRAYKPPAQTSPAGSPREPPLOCLVLI 267  
 QY 238 EAIPLPFDGA 249  
 DB 268 EAIPLPFDGA 275

## RESULT 10

AAB34826  
 ID AAB34826 standard; protein; 705 AA.

XX AAB34826;

XX 28-MAY-2003 (first entry)

XX Human hypoxia inducible factor (HIF)-3alpha protein.

XX Human; hypoxia inducible factor; HIF-3alpha; tumour; muscular function;  
 XX vascular endothelial growth factor; VEGF; angiogenesis; cardiovascular;  
 XX coronary disease; cardiac disease; cytostatic; cardiac; cell survival.

XX Homo sapiens.

XX MO20294862-A2.

XX 28-NOV-2002.

XX 23-MAY-2002; 2002MO-CA000752.

XX 23-MAY-2001; 2001US-0292630P.

XX 08-FEB-2002; 2002US-0354529P.

XX (ANGI-) ANGIOGENE INC.

XX Guy L;

XX WPI; 2003-129410/12.

XX N-PSDB; AAD53131.

XX New Hypoxia Inducible Factor-3 alpha polynucleotide for inducing vascular  
 XX endothelial growth factor (VEGF) expressions in a tissue or cell,  
 XX particularly for eliminating tumors, or treating coronary or cardiac

PT diseases in mammals.

XX Claim 13; Col 69-72; 38pp; English.

XX The invention relates to hypoxia inducible factor (HIF)-3alpha DNA and  
 CC protein sequences. The HIF-3alpha DNA and protein sequences are useful  
 CC for inducing vascular endothelial growth factor (VEGF) expression or  
 CC angiogenesis in a mammalian tissue or cell, for modulating tumoural cell  
 CC survival, eliminating a tumoural cell, or evaluating the malignancy of a  
 CC tumour in a subject. They are also useful for improving muscular  
 CC function, particularly for treating coronary and cardiac diseases in  
 CC mammals. The present sequence is human HIF-3alpha protein

XX  
 SQ Sequence 705 AA;

Query Match 48.6%; Score 795; DB 6; Length 705;  
 Best Local Similarity 67.6%; Pred. No. 8.6e-73;  
 Matches 171; Conservative 11; Mismatches 29; Indels 42; Gaps 5;

QY 9 RSNTLEKREKSRDAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLMHR 68  
 DB 3 STTLEKREKSRDAARSRSQETEVLYQLAHTLPFARGVSAHLDKASIMRLTISYLMHR 62  
 QY 69 LCAAAGKRGATGRLLPEGGGFRHGTNRGRHGLPVGKCQ-----QAPGQSV 117  
 DB 63 LCAAAG-----EWNQVAGGGEPLDACYKALBGFVWVLTAEQDMAY 102  
 QY 118 LCSSSLIHNPTPGTNFS-LELIGHSTFDFIHPDQDELDALTPRPNLSSKKLEAPTERH 176  
 DB 103 -----LSENVSKHLGLSQLELIGHSTFDFIHPDQDELDALTPQQLTSRRKVEAPTER 156  
 QY 177 FSLRMKSTLTSGRTLNKAAATWVKVLCGSHMRAYKPPAQTSPAGSPREPPLOCLVLI 236  
 DB 157 FSLRMKSTLTSGRTLNKAAATWVKVLCGSHMRAYKPPAQTSPAGSPREPPLOCLVLI 216  
 QY 237 CEAIPLPFDGA 249  
 DB 217 CEAIPLPFDGA 225

## RESULT 11

AAB94934  
 ID AAB94934 standard; protein; 648 AA.

XX AAB94934;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:16415.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUN-2000; 2000EP-00116126.

XX 29-JUN-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Oca T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX Iehi S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.

PT Primer sets for synthesising polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

PS Claim 8; SEQ ID NO 16415; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AA982446 to AB958939 represent human amino acid sequences, and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

SQ Sequence 648 AA;

Query Match	35.0%;	Score 573;	DB 4;	Length 648;
Best Local Similarity	61.0%;	Pred. No. 7.7e-50;		
Matches 125; Conservative	11;	Mismatches 27;	Indels 42;	Gaps 5;

QY 57 MRLTISYLMRLCAAGKGRATGRLPEGPGGFRHGTNRGRHGLPVGKQ----- 109

Db 1 MRLTISLRMRLCAAG-----EWNQVAGGGEPLDACYLKALEGF 40

110 ---QAPGQSVDLCCSSLIHNPYPGTNFS-LEIGHSIFDEIHPCDQEELODALTPRPN 164

Db 41 VMVLTAEQDMAY-----LSENVSKHLGLSQLELIGHSIFDFIHPCDQEELQDALTPQQT 94

QY 165 LSKKLEAPTERHFSLRMKS TLTSRGRTLN LKATWV LHCSGHMRAYKPPAQTSPAGSP 224

Db 95 LSRKVEPTERCFSLRKSTLTSGRTLNKATWVNLNCSGMRAYKPPAQSPAGSP 154

QY 225 RSEPLQCLVLICCAIPQLPFHDGA 249

Db 155 DSEPPLOCLVLICAI P----HPGS 175

RESULT 12  
AAB93710  
ID AAB93710 standard; protein; 199 AA

AC AAB93710;

DT 26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:13303.

**KW** Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00248036.

PR 11-JAN-2000; 2000JP-00118776.

PR 09-JUN-2000; 2000JP-00241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa

XX

1000

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full  
PT length cDNAs defined in the specification, and for the detection and/or  
PT diagnosis of the abnormality of the proteins encoded by the full-length  
PT cDNAs.

PS Claim 8; SEQ ID NO 13303; 2537pp + Sequence Listing; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH932446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 199 AA;

Query Match	31.7%	Score 518.5;	DB 4;	Length 199;
Best Local Similarity	68.0%;	Pred. No. 6.2e-45;		
Matches 104; Conservative	9;	Mismatches 13;	Indels 27;	Gaps 21

QY 120 CSSSLHNPTGTF-----SLEIGHSIFDFIHPCDQEELO 156

Db 14 CTSWLT RCPSPASAP TWT RPLSCASPSATCACTASAPQLELIGHSI FDFIHPCDQEEIQ 73

QY 157 DALTPRNLKKLEAPTERHFSLRMKSTJTSRGRTLNLKAATWKLHCSGIMRAYKPPA 216

Db 74 DALTPQQLSRKVEAPTERCFSLRMKSTLTSGRTLNLKATWKLNCSGHMRAYKPPA 133

217 QTSPAGSPREPLQCLVLCIAIPQLPFHDGA 249

Db 134 QTSPAGSPDSEPLQLCLVLCIAIP---HPGS 162

RESULT 13  
AD16229  
ID AD16229 standard: protein: 513 AA

XX  
AC  
AD116229:

XX  
DT 22-APR-2004 (first entry)

XX Human nucleic acid-associated protein (NAAP) #14  
DE

XX human: nucleic acid-associated protein; NAAP: autoimmune disorder;



KM	inflammatory disorder; AIDS; allergy; infection; metabolic disorder;
KM	obesity; reproductive disorder; infertility; neurological disorder;
KM	Parkinson's disease; Alzheimer's disease; cardiovascular disorder;
KM	myocardial infarction; hypertension; eye disorder;
KM	cell proliferative disease; cancer.
XX	
OS	Homo sapiens.
XX	
PN	WO2003094848-A2.
XX	
PD	20-NOV-2003.
XX	
PF	09-MAY-2003; 2003WO-US014450.
XX	
PR	10-MAY-2002; 2002US-0379843P.
PR	24-MAY-2002; 2002US-0383457P.
PR	31-MAY-2002; 2002US-0384699P.
PR	06-JUN-2002; 2002US-0387265P.
XX	
PA	(INCY-) INCYTE CORP.

PI Kable AE, Elliott VS, Tian UK, Ramkumar J, Margolis JF, Chawla NK,  
PI Richardson TW, Bullock SA, Khare R, Lee SY, Lal PG, Tang YT, Yue H,  
PI Swernakar A, Becha SO, Hafeilia AJA, Chang H, Baugh MR, Borowsky M,  
PI Gleason KJ, He A, Foryshtie IU, Sprague WN, Blake JN, Warren BA,  
PI Mason PM, Ison CH, Lindquist EA, Wilson AD, Jin P,  
XX  
WRP: 2004-011999/01.  
DR  
N-PSDBJ: AD116279.

PT New human nucleic acid associated proteins and polynucleotides, useful for diagnosing, preventing or treating diseases or conditions associated with aberrant protein expression, e.g. cancer, AIDS, atherosclerosis or stroke.

PS Claim 1; SEQ ID NO 14; 400pp; English

The invention comprises the amino acid and coding sequences of human nucleic acid-associated proteins (NAAP). The DNA and protein sequences of the invention are useful in diagnosing, preventing and treating diseases/conditions associated with altered expression of NAAP, such as: autoimmune/inflammatory disorders (e.g. AIDS and allergies), infections (e.g. bacterial and viral), metabolic disorders (e.g. obesity), reproductive disorders (e.g. infertility), neurological disorders (e.g. Parkinson's disease and Alzheimer's disease), cardiovascular disorders (e.g. myocardial infarction and hypertension), eye disorders, or cell proliferative diseases (e.g. cancer). The present amino acid sequence represents a human NAAP protein of the invention.

**SQ Sequence 513 AA;**

Query Match	30.4%	Score 498	DB 8	Length 513
Best Local Similarity	46.2%	Pred. NO. 3.2e-42		
Matches 115	Conservative 26	Mismatches 60	Indels 48	Gaps 6

```
QY      9 RSNTELRKEKSRDAARSRRSOETVELYLQAHTLPARGVSAHLDKASIMPLTISYLRMR 688
        ::::::::::::::::::::: : ::::::::::::::::::::: :
DB     13 RISSERKRKEKSRDARSRSKESEVFYELAHOQLPHNNVSSHLDKASVMPLTISYLRVRK 722
```

69 LCAAGGKGRATGRLPEGPGGFHGTNRGRHGLPVGKQQA-----PGPQSV 118

Db 73 LLDAG-----DLDIEDDMKAQMNCFYLLKALDGFVVL 104

119 LCSSSLH-----NPTPG-TNFSLELIGHSIFDFIHPCDQELQDATTPRPNLSKKLEA 1722

Db 105 TDDGDMIIYISDNVNMGLTQF--ELTGHSVFDFTHPCDHEEMREMLTHRNGLVKKKEQ 162

173 PTERHSLRMKSTLTSRGRTLNKATWVLHCSGIMRAYKPPAQTSAGSPRSEPLQC 232

Db 163 NTQRSFFLRMKCTLTSGRTMNIKSATWKVLHCTGHIHVYDTNSNQPGG--YKKPMT C 220

QY	233	LV	LICE	AI	P	241
----	-----	----	------	----	---	-----

Db 221 LVLICEPIR 229

RESULT 14  
ADE25733  
ID ADE25733 standard; protein; 827 AA

XX 29-JAN-2004 (first entry)  
DT

DE Human protein differentially expressed in foam cells #10.

KM Human; differential expression; foam cell; LPS; lipopolysaccharide  
KM cardiovascular disease; atherosclerosis.

**OS Homo sapiens.**

PN US2003194721-A1.

PD 16-OCT-2003.

PF 18-SEP-2002; 2002US-00247671.

PR 19-SEP-2001; 2001US-0323784P.

PA (INCY-) INCYTE GENOMICS INC.

PI Mikita T, Shiffman D, Porter JG, Kaber MR;

DR WPI; 2003-875398/81.

PT Combination containing several polynucleotide that are differentially  
PT expressed in foam cells and complements of the polynucleotides, useful  
PT for diagnosing cardiovascular disease or atherosclerosis.

PS Disclosure; SEQ ID NO 137; 37pp; English.

The invention relates to a combination comprising several polynucleotides having any one of 127 sequences (S1) such as the sequence of human calmodulin gene, human mRNA for KIAA0930 protein, leukotriene A<sub>4</sub> hydrolase, human CG1-142 protein mRNA, human K<sup>+</sup> channel beat 2 subunit mRNA, etc., and their complements. The cDNAs are differentially expressed in LPS (lipopolysaccharide)-treated foam cells. Also included are obtaining an extended or full length gene from a library of nucleic acid sequences, an expression vector containing the nucleic acids, a host cell containing the vector, a purified polypeptide appearing as ADB25750 and ADB25751, producing a protein by culturing the host cell, and a composition comprising a purified antibody that specifically binds to the proteins. The foam cell-expressed nucleic acids are useful for a high throughput detection of differential expression of one or more polynucleotides in a sample. The sample is from a subject with atherosclerosis and comparison with a standard defines early, mid or late stages of the disorder. The foam cell-expressed nucleic acids are useful for high throughput screening of a library of molecules or compounds to identify a ligand which binds a polynucleotide. The library is chosen from DNA molecules, peptides, proteins and RNA molecules. The protein is useful for a high throughput screening of library of molecules or compounds to identify at least one ligand which specifically binds a protein, for purifying a ligand from a sample for making an antibody. The foam cell-expressed nucleic acids are useful for diagnosing cardiovascular disorder. The foam cell-expressed nucleic acids are useful as elements on a microarray which can be used for detecting related polynucleotide in a sample, diagnosing cardiovascular disease atherosclerosis. The present sequence represents a protein differentially expressed in LPS treated foam cells.

**SQ** Sequence 827 AA;

Query Match	30.4%	Score 498;	DB 7;	Length 827;
Best Local Similarly	46.2%	Pred. No. 6.5e-42;		
Matches 115; Conservative	26;	Mismatches 60;	Indels 48;	Gaps 6

QY 9 RSTNLEKESRDAAARRSQETVLYQLAHTLPPRAGVSAHLDKASIMRLTISYLRMR 68  
 DB 13 RISSEKESRDAAARRSQETVLYQLAHTLPPRAGVSAHLDKASIMRLTISYLRMR 72  
 QY 69 LCAAGGKRGATGRLLPEGPGFRHGTNRGRHGLPVGKCGQA-----PGQSYD 118  
 DB 73 LLDNG-----DLIDIDMKAMQNCIFYLKALDGFVWVL 104  
 QY 119 LCSSSLIH-----NPTPG-TNFSLELIGHSIFDFIHPDQDELDALTPRPNLKXKLEA 172  
 DB 105 TDDGDMIVISDNVKNKMGTLQF--ELTGHSVFDFTHPCDHEMRMLTHRNGLVKKGKEQ 162  
 QY 173 PTERHFSLRKMSLTLSRGRTLNKAATWKLHSGHMRAYKPPAQTSPPAGSPRSEPPLOC 232  
 DB 163 NTRRSFPLRMKCTLSRGRTNKAATWKLHSGHMRAYKPPAQTSPPAGSPRSEPPLOC 220  
 QY 233 LVLICEAIP 241  
 DB 221 LVLICEPIP 229

## RESULT 15

AD76984 standard; protein; 827 AA.

29-JAN-2004 (first entry)

Human protein expressed in a liver disorder #43.

human; liver disorder; hyperlipidaemia; hypertension; type II diabetes;  
 tumour; liver; inflammatory disorder; immune response disorder;  
 high-throughput screening; differential gene expression; gene therapy.

Homo sapiens.

US2003108871-A1.

12-JUN-2003.

30-JUL-2001; 2001US-00919039.

28-JUL-2000; 2000US-0222113P.

(KASE/) KASER M R.

Kaser MR;

WPI, 2004-031227/03.

N-PSDB; AD76983.

Composition comprising several cDNAs that are differentially expressed in  
 treated human CJA liver cell cultures, useful for treating liver  
 disorders.

Claim 1; SEQ ID NO 149; 41pp; English.

The invention relates to a composition comprising several cDNAs that are  
 differentially expressed in a liver disorder. The composition is useful  
 for treating liver disorder such as hyperlipidaemia, hypertension, type  
 II diabetes, tumours of the liver and disorders of the inflammatory and  
 immune response. The composition is useful for a high-throughput method  
 of screening several molecules or compounds to identify a ligand which  
 specifically binds a cDNA. A protein encoded by the cDNA is useful for a  
 high-throughput method for using a protein to screen several molecules or  
 compounds to identify at least one ligand which specifically binds the  
 protein which involves combining the protein encoded by the cDNA with  
 several of molecules or compounds under conditions to allow specific  
 binding, and detecting specific binding between the protein and a  
 molecule or compound, therefore identifying a ligand which specifically  
 binds the protein. The composition is useful for detecting and

quantifying differential gene expression, can be used in gene therapy, to  
 formulate prognosis and to design a treatment regimen and to monitor the  
 efficacy of treatment. The present sequence represents the amino acid  
 sequence of a protein encoded by a cDNA differentially expressed in a  
 liver disorder.

Sequence 827 AA;

Query Match 30.4%; Score 498; DB 8; Length 827;

Best Local Similarity 46.2%; Pred. No. 6.5e-42;

Matches 115; Conservative 26; Mismatches 60; Indels 48; Gaps 6;

QY 9 RSTNLEKESRDAAARRSQETVLYQLAHTLPPRAGVSAHLDKASIMRLTISYLRMR 68  
 DB 13 RISSEKESRDAAARRSQETVLYQLAHTLPPRAGVSAHLDKASIMRLTISYLRMR 72  
 QY 69 LCAAGGKRGATGRLLPEGPGFRHGTNRGRHGLPVGKCGQA-----PGQSYD 118  
 DB 73 LLDNG-----DLIDIDMKAMQNCIFYLKALDGFVWVL 104  
 QY 119 LCSSSLIH-----NPTPG-TNFSLELIGHSIFDFIHPDQDELDALTPRPNLKXKLEA 172  
 DB 105 TDDGDMIVISDNVKNKMGTLQF--ELTGHSVFDFTHPCDHEMRMLTHRNGLVKKGKEQ 162  
 QY 173 PTERHFSLRKMSLTLSRGRTLNKAATWKLHSGHMRAYKPPAQTSPPAGSPRSEPPLOC 232  
 DB 163 NTRRSFPLRMKCTLSRGRTNKAATWKLHSGHMRAYKPPAQTSPPAGSPRSEPPLOC 220  
 QY 233 LVLICEAIP 241  
 DB 221 LVLICEPIP 229

Search completed: June 15, 2005, 15:57:10  
 Job time : 165 secs